

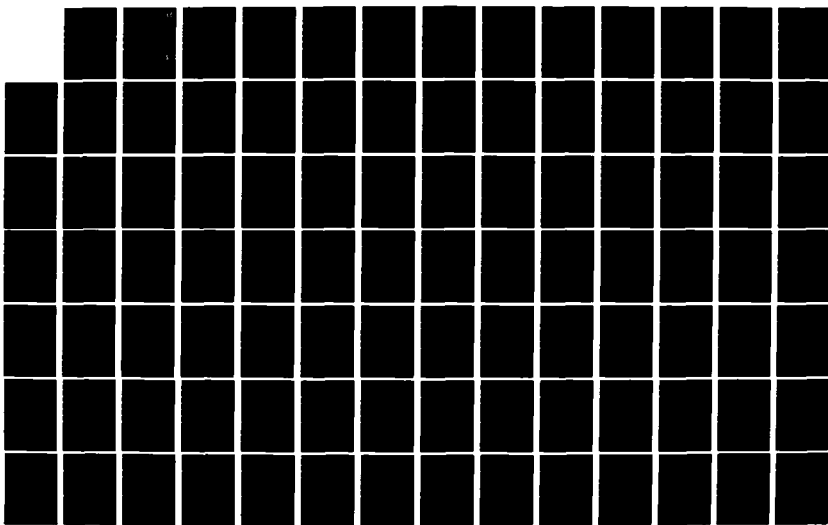
AD-A124 975

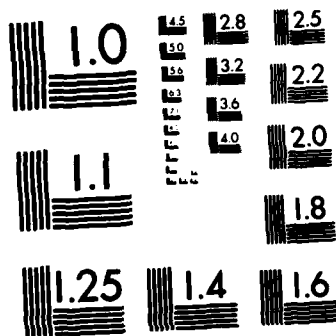
INFORMATION THEORETIC DETECTION OF OBJECTS EMBEDDED IN  
CLUTTERED AERIAL SCENES(U) AIR FORCE INST OF TECH  
WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI... F D TILLER  
JUL 82 AFIT/DS/EE/82-1 .F/G 1475

1/2

UNCLASSIFIED

NL





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

AD A124975



①

INFORMATION THEORETIC DETECTION  
OF OBJECTS EMBEDDED IN  
CLUTTERED AERIAL SCENES

DISSERTATION

AFIT/DS/EE/82-1 Frank D. Tiller  
Captain USAF

DTIC  
ELECTE

FEB 25 1983

S

D

DTIC FILE COPY

UNITED STATES AIR FORCE  
AIR UNIVERSITY  
AIR FORCE INSTITUTE OF TECHNOLOGY  
Wright-Patterson Air Force Base, Ohio

83 02 024 029

**DISTRIBUTION STATEMENT A**

Approved for public release  
Distribution Unlimited

AFIT/DS/EE/82-1

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



INFORMATION THEORETIC DETECTION  
OF OBJECTS EMBEDDED IN  
CLUTTERED AERIAL SCENES

DISSERTATION

AFIT/DS/EE/82-1 Frank D. Tiller  
Captain USAF

Approved for public release; distribution unlimited

AFIT/DS/EE/82-1

INFORMATION THEORETIC DETECTION  
OF OBJECTS EMBEDDED IN  
CLUTTERED AERIAL SCENES

DISSERTATION

Presented to the Faculty of the School of Engineering  
of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the  
Requirements for the Degree of  
Doctor of Philosophy

by

Frank D. Tiller, B.S., M.S.

Captain

USAF

July 1982

Approved for public release; distribution unlimited

AFIT/TS/EE/82-1

INFORMATION THEORETIC DETECTION OF OBJECTS  
EMBEDDED IN CLUTTERED AERIAL SCENES

by

Frank D. Tiller, B.S., M.S.  
Captain USAF

Approved:

Matthew Krosky  
Chairman

19 July 1982

Donald J. Cygnar

19 July 1982

Robert S. Moore

19 July, 1982

Dennis W. Zelnin

19 JULY, 1982

Richard W. Kulp

19 July, 1982

Accepted:

JSPremieniachi  
Dean, School of Engineering

20 July 1982

## ACKNOWLEDGEMENTS

Many people have helped provide the foundation for this work, too many to mention individually, however, I owe a special debt to the members of my advisory committee for their enthusiastic support and advice throughout the course of this research. I especially want to thank Dr. Matthew Kabrisky, my committee chairman, for providing the advice, stimulating atmosphere and encouragement required to complete this dissertation. Finally, to my wife, Jackie, and to my daughter, Nicole, I wish to express my love and gratitude for their patience and understanding during the entire graduate program.

## Contents

	Page
Acknowledgements . . . . .	iii
List of Figures . . . . .	vi
List of Tables . . . . .	x
Abstract . . . . .	xi
I. Introduction . . . . .	1
II. Maximum Entropy and Minimum Cross-Entropy . . .	9
Background . . . . .	9
Definitions and General Problem Statement .	11
The Consistency Axioms . . . . .	13
Properties of Cross-Entropy Minimization .	16
III. Detection Algorithm Development . . . . .	25
Introduction . . . . .	25
Solution of the Constraint Equations . . .	26
Solution Characteristics . . . . .	30
Further Minimum Cross-Entropy Properties .	31
The Detection Algorithm . . . . .	36
IV. Information Functions . . . . .	47
Introduction . . . . .	47
Image Moments . . . . .	47
Orthonormal Moments . . . . .	47
Legendre Polynomials . . . . .	61
V. Numerical Techniques and Performance Analysis .	65
Lambda Vector Solution . . . . .	67
Numerical Quadrature . . . . .	70



	Page
Detection Algorithm Infrastructure . . . . .	74
Performance Analysis . . . . .	76
VI. Processing Results . . . . .	86
Performance Factors . . . . .	86
Approximate Error Probability . . . . .	91
VII. Summary and Future Research . . . . .	117
Summary . . . . .	117
Future Research . . . . .	119
Bibliography . . . . .	123
Appendix A: Template Densities . . . . .	128
Appendix B: Minimum Cross-Entropy Templates . . . . .	147
Appendix C: Template Photographs . . . . .	166

## List of Figures

Figure	Page
1.1 Generation of a Vertical Aerial Scene . . . . .	3
3.1 Posterior Adaptation . . . . .	37
3.2 Density Function Evolution . . . . .	41
4.1 Image Ellipse . . . . .	52
4.2 Alternate Methods to Compute Invariant Moments . . . . .	55
4.3 Information Function Construction . . . . .	59
5.1 Software Module Interdependence . . . . .	75
5.2 Probability of Error versus Correlation Difference over Sigma . . . . .	82
6.1 Template to Scene Cross-Entropy versus Approximation Order . . . . .	88
6.2 Scene Cross-Entropy versus Scene/Target Ratio . . . . .	90
6.3 Approximate Template to Scene Cross-Entropy . .	92
6.4 Block Tank Density . . . . .	93
6.5 Minimum Cross-Entropy Tank Reconstruction (6 Information Functions) . . . . .	94
6.6 Minimum Cross-Entropy Tank Reconstruction (10 Information Functions) . . . . .	95
6.7 Minimum Cross-Entropy Tank Reconstruction (15 Information Functions) . . . . .	96
6.8 Minimum Cross-Entropy Tank Reconstruction (21 Information Functions) . . . . .	97
6.9 Minimum Cross-Entropy Tank Reconstruction (28 Information Functions) . . . . .	98
6.10 Minimum Cross-Entropy Tank Reconstruction (36 Information Functions) . . . . .	99

Figure		Page
6.11	Minimum Cross-Entropy Tank Reconstruction (45 Information Functions) . . . . .	100
6.12	Minimum Cross-Entropy Tank Reconstruction (55 Information Functions) . . . . .	101
6.13	Minimum Cross-Entropy Tank Reconstruction (66 Information Functions) . . . . .	102
6.14	Minimum Cross-Entropy Tank Reconstruction (78 Information Functions) . . . . .	103
6.15	Minimum Cross-Entropy Tank Reconstruction (91 Information Functions) . . . . .	104
6.16	Test Scene with Tank and Clutter . . . . .	107
6.17	Test Scene with Clutter . . . . .	108
6.18	Probability of Error versus Template Alternatives . . . . .	112
6.19	Probability of Error versus Cross-Entropy. . . . .	115
A.1	Template Density 1 with Cross-Entropy = 0.71460 . . . . .	129
A.2	Template Density 2 with Cross-Entropy = 0.86036 . . . . .	130
A.3	Template Density 3 with Cross-Entropy = 1.29606 . . . . .	131
A.4	Template Density 4 with Cross-Entropy = 0.93407 . . . . .	132
A.5	Template Density 5 with Cross-Entropy = 0.79444 . . . . .	133
A.6	Template Density 6 with Cross-Entropy = 0.90418 . . . . .	134
A.7	Template Density 7 with Cross-Entropy = 0.87136 . . . . .	135
A.8	Template Density 8 with Cross-Entropy = 0.96869 . . . . .	136
A.9	Template Density 9 with Cross-Entropy = 0.86579 . . . . .	137

Figure		Page
A.10	Template Density 10 with Cross-Entropy = 0.91397 . . . . .	138
A.11	Template Density 11 with Cross-Entropy = 0.85447 . . . . .	139
A.12	Template Density 12 with Cross-Entropy = 0.92389 . . . . .	140
A.13	Template Density 13 with Cross-Entropy = 0.88835 . . . . .	141
A.14	Template Density 14 with Cross-Entropy = 0.91680 . . . . .	142
A.15	Template Density 15 with Cross-Entropy = 0.88692 . . . . .	143
A.16	Template Density 16 with Cross-Entropy = 0.91118 . . . . .	144
A.17	Template Density 17 with Cross-Entropy = 0.91693 . . . . .	145
A.18	Template Density 18 with Cross-Entropy = 0.88837 . . . . .	146
B.1	Minimum Cross-Entropy Template 1 with Cross-Entropy = 0.34217 . . . . .	148
B.2	Minimum Cross-Entropy Template 2 with Cross-Entropy = 0.40756 . . . . .	149
B.3	Minimum Cross-Entropy Template 3 with Cross-Entropy = 0.44235 . . . . .	150
B.4	Minimum Cross-Entropy Template 4 with Cross-Entropy = 0.44962 . . . . .	151
B.5	Minimum Cross-Entropy Template 5 with Cross-Entropy = 0.41877 . . . . .	152
B.6	Minimum Cross-Entropy Template 6 with Cross-Entropy = 0.44274 . . . . .	153
B.7	Minimum Cross-Entropy Template 7 with Cross-Entropy = 0.4446 . . . . .	154
B.8	Minimum Cross-Entropy Template 8 with Cross-Entropy = 0.46145 . . . . .	155

Figure		Page
B.9	Minimum Cross-Entropy Template 9 with Cross-Entropy = 0.42566 . . . . .	156
B.10	Minimum Cross-Entropy Template 10 with Cross-Entropy = 0.46862 . . . . .	157
B.11	Minimum Cross-Entropy Template 11 with Cross-Entropy = 0.43961 . . . . .	158
B.12	Minimum Cross-Entropy Template 12 with Cross-Entropy = 0.51369 . . . . .	159
B.13	Minimum Cross-Entropy Template 13 with Cross-Entropy = 0.39674 . . . . .	160
B.14	Minimum Cross-Entropy Template 14 with Cross-Entropy = 0.42777 . . . . .	161
B.15	Minimum Cross-Entropy Template 15 with Cross-Entropy = 0.39324 . . . . .	162
B.16	Minimum Cross-Entropy Template 16 with Cross-Entropy = 0.41373 . . . . .	163
B.17	Minimum Cross-Entropy Template 17 with Cross-Entropy = 0.44381 . . . . .	164
B.18	Minimum Cross-Entropy Template 18 with Cross-Entropy = 0.41452 . . . . .	165
C.1	Template Alternative One . . . . .	167
C.2	Template Alternative Two . . . . .	168
C.3	Template Alternative Three . . . . .	169
C.4	Template Alternative Four . . . . .	170
C.5	Template Alternative Five . . . . .	171
C.6	Template Alternative Six . . . . .	172
C.7	Template Alternative Seven . . . . .	173
C.8	Template Alternative Eight . . . . .	174
C.9	Template Alternative Nine . . . . .	175

List of Tables

Table	Page
I    Legendre Polynomials . . . . .	63

Abstract

This dissertation presents the theoretical development and numerical implementation of a minimum cross-entropy target detection algorithm. The procedure is based on the solution of a nonlinear constrained cross-entropy minimization problem and requires information in the form of raw image moments. The detection rule involves both preprocessing and real-time computations. The preprocessing requires the selection of a set of target templates and the solution of the constrained cross-entropy minimization problem for the selected target templates. The real-time processing requires the computation of image moments and a set of dot product operations.

An orthonormal set of "information functions" is developed and numerical methods of converting raw image moments into the expected values of the information functions are given. Numerical techniques for image moment computation and a solution scheme for the nonlinear set of constraints are developed and implemented. The theoretical development of the detection algorithm is given starting from a set of consistency axioms. The expected performance is analyzed and factors determining performance presented. The procedure is applied to a test set of 100 images and the detection algorithm error probability is projected and related to the salient performance determining factors.

# INFORMATION THEORETIC DETECTION OF OBJECTS EMBEDDED IN CLUTTERED AERIAL SCENES

## Chapter I. Introduction

The general problem considered in this dissertation is that of characterizing and evaluating the information present in the image plane of an optical system. The specific problem of interest is the detection of complex man-made objects in aerial scenes that contain confusing background information or optical clutter. A general overview of this target detection in clutter problem can be found in Gagnon's dissertation (Gagnon, 1975) while Harley et al (Harley, 1977) provide an overview of typical system parameters encountered in practice.

The image plane which is the source of information for this detection problem is usually a photograph which can be taken from any airborne vehicle. The source of information or aerial photographs are classified as either vertical or oblique aerial photographs depending on the angle of inclination of the optical axis of the lens. Vertical photographs are those taken with the optical axis of the lens pointing vertically downward at the time of exposure. Oblique photographs are those taken with the optical axis intentionally deviated from the vertical. Oblique photographs are further classified as low and high oblique based on the magnitude of



the angle of deviation. A low oblique has a relatively small or low angle of deviation from vertical and does not include the apparent horizon or the visible junction of earth and sky as seen from the camera station. A high oblique has a relatively large or high angle of deviation from the vertical and includes the apparent horizon (Whitmore, 1966: 1). This dissertation will only characterize vertical photographs taken from a known altitude, however, the methods used in this work should also characterize at least low oblique photographs.

Figure 1.1 shows some of the geometry involved in generating a vertical aerial photograph. Each camera exposure produces a frame of information that is shown as a series of large non-overlapping squares in the figure for simplicity. The frames are also shown partitioned into  $K^2$  "information cells" that form the basic decision elements for the detection algorithm. The objects to be located belong to one of a set of known classes and all elements in a given class are essentially identical. The class of objects of current interest is called the target and a target can appear at any location and orientation within a frame.

With this problem formulation the only information available for target detection is the image plane irradiance distribution function  $I(x,y)$  that is the image of the cluttered ground scene. Formation of this image  $I(x,y)$  from an object scene  $F(\xi,\eta)$  actually represents a flow of in-

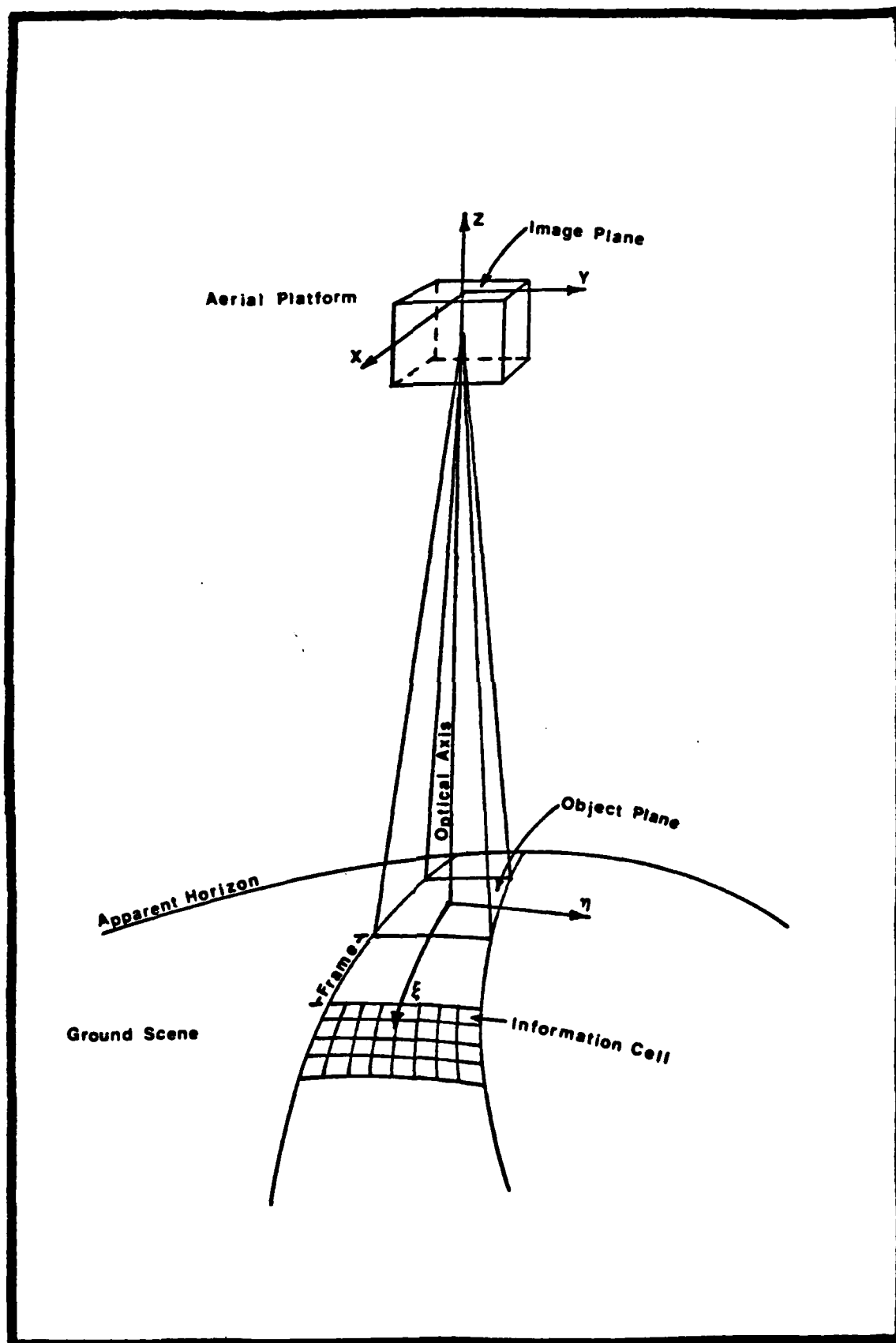


Fig. 1.1 Generation of a Vertical Aerial Scene

formation from the object plane to the image plane. The carriers of information are the photons. In traveling from the object plane to the image plane, a photon encounters intervening physical processes such as lenses and the atmosphere. The sum of these processes forms an information channel. In its most basic sense, the irradiance distribution function is no more than a superposition of photon events. These events are photon arrivals (units of irradiance) for the image or photon departures (units of radiance) from the object. The sum or density of the photon events as a function of position in the image plane defines the irradiance distribution function (Frieden, 1979). It is this irradiance distribution function  $I(x,y)$  or photon density that represents the information channel output and that must be characterized and used in the detection algorithm.

Looking at the image more mathematically, let  $C(x,y,t,\lambda)$  represent the spatial energy distribution of an image source of radiant energy at spatial coordinates  $(x,y)$  at time  $t$  and a wavelength  $\lambda$ . Because light intensity is a real positive quantity, that is because intensity is proportional to the modulus squared of the electric field, the image light function is real and non-negative. Furthermore, in all practical imaging systems there is always a small amount of background light present. Because of this background light and the physical restrictions imposed by the imaging system, it is assumed that

$$0 < C(x,y,t,\lambda) \leq A$$

where  $A$  is the maximum brightness. An image is also necessarily limited in extent by the imaging system and the recording media. For mathematical simplicity all images in this dissertation are assumed to be nonzero only over a square region for which

$$-L \leq x,y \leq L$$

Since the image is also observable only for a finite time,  $(-T \leq t \leq T)$  the image light function  $C(x,y,t,\lambda)$  is a bounded four-dimensional function with bounded independent variables. As a final restriction, it is assumed that the image light function is continuous over its domain of definition (Pratt, 1978: 4). The image light function  $C(x,y,t,\lambda)$  is actually at worst piece-wise continuous and is well approximated by a continuous function.

The brightness response to the image light function  $C(x,y,t,\lambda)$  can now be defined for both men and machines. In men the brightness response of a standard human observer is commonly used to define the instantaneous luminance of the light field as shown by

$$Y(x,y,t) = \int_0^{\infty} C(x,y,t,\lambda) V_s(\lambda) d\lambda$$

where  $V_s(\lambda)$  represents the relative luminous efficiency function or the spectral response of human vision. Similarly, the color response of a standard human observer is measured and used in terms of some set of tristimulus values that are linearly proportional to the amounts of red, green, and blue light needed to "match" a colored light. In a machine with a multispectral imaging system the observed image field is modeled as a spectrally weighted integral of the image light function. The  $i$ th spectral image field is then given by

$$F_i(x,y,t) = \int_0^{\infty} C(x,y,t,\lambda) S_i(\lambda) d\lambda$$

where  $S_i(\lambda)$  is the spectral response of the  $i$ th sensor (Hall, C., 1978: 17). For a monochrome imaging system, as will be used in this dissertation, the image function  $F(x,y,t)$  nominally denotes the image luminance or some converted or corrupted physical representation of luminance.

The image function  $F(x,y,t)$  is propagated through the information channel or transformed from the object scene plane to the image plane of the aircraft to form the instantaneous irradiance distribution. The channel transformation can be viewed as a one-to-one mapping and is defined by

$$I(x,y,t) = T\{F(x,y,t)\}$$

When the transformation is also assumed to be an additive

linear operator the standard superposition integral description of the channel output is obtained. Using the sifting property of delta functions the mapping is first rewritten as

$$I(x,y,t) = T \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\xi,\eta,t) \delta(x-\xi,y-\eta) d\xi d\eta \right\}$$

Now changing the order of the general linear operator  $T$  and the integral operator results in the expression

$$I(x,y,t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\xi,\eta,t) T \left\{ \delta(x-\xi,y-\eta) \right\} d\xi d\eta$$

Then defining the channel point spread function as  $H(x,y;\xi,\eta) = T \left\{ \delta(x-\xi,y-\eta) \right\}$  gives the desired integral expression for the channel output  $I(x,y,t)$ . The superposition integral description of the channel output or irradiance distribution is given by

$$I(x,y,t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\xi,\eta,t) H(x,y;\xi,\eta) d\xi d\eta$$

In the object detection problem of interest in this work, the image does not change with time and the time variable can be dropped from the instantaneous irradiance distribution to let  $I(x,y)$  represent the spatial distribution of light in the image plane or the light density function.

Now the normalized irradiance distribution can be defined by

$$i(x,y) = I(x,y) / \int_{-L}^L \int_{-L}^L I(x,y) dx dy$$

The normalized distribution function has all the properties of a bivariate probability density function since

$$\int_{-L}^L \int_{-L}^L i(x,y) dx dy = 1$$

and the probability of a photon arriving in any region  $R$  of the image is given by the expression

$$P(R) = \iint_R i(x,y) dx dy$$

Several other authors have used this probability density viewpoint in their work in image processing. Among these are Frieden, working with image restoration techniques (Frieden, 1972) and Minerbo, in reconstructing a source from a discrete set of projection data (Minerbo, 1979). Using this viewpoint it is this bivariate probability density function that must be characterized and used in the detection algorithm.

## Chapter II. Maximum Entropy and Minimum Cross-Entropy

The principles of maximum entropy and minimum cross-entropy provide a means of approximating the normalized irradiance distribution  $i(x,y)$  and detecting targets of interest in a cluttered aerial scene. The approach taken in developing a detection rule for the cluttered scene problem of this dissertation is based on entropy and cross-entropy having unique properties as information measures (Johnson, 1979) and on cross-entropy minimization having unique properties as an inference procedure (Shore, 1980). The work is an extension of Miller's work (Miller, 1980) in approximating one-dimensional probability density functions using a maximum entropy criterion and much of the background material is reviewed in his dissertation.

### Background

The principles of entropy maximization and cross-entropy minimization both have their roots in Shannon's work in communication theory. For discrete, noiseless systems, maximizing the source entropy results in the best source encoding in the sense of enabling the highest information rate over a fixed capacity channel (Shannon, 1948a). For continuous systems, Shannon's definition of source rate for a fixed fidelity criterion or rate-distortion function involved the mini-



mization of a functional (mutual information) like cross-entropy (Shannon, 1948b). However, it was Edwin Jaynes who first proposed the principle of maximum entropy as a means of approximating an unknown probability density function more than twenty-five years ago (Jaynes, 1957). While the name cross-entropy is due to Good (Good, 1963), the principle of minimum cross-entropy is a generalization of the maximum entropy principle that was first proposed by Kullback, who called it a principle of minimum directed divergence or minimum discrimination information (Kullback, 1959:37). Jaynes' work has been applied in a number of areas, but within the engineering community the most widely known application is Burg's Maximum Entropy Spectral Analysis (MESA) technique (Burg, 1967). However, the maximum entropy principle is applied indirectly in terms of filtering, rather than directly in terms of approximating the underlying probability densities and it is not widely understood that MESA is identical to Jaynes' principle (Shore, 1981). Despite their many proven applications, Jaynes' principle of maximum entropy and Kullback's principle of minimum cross-entropy have had a controversial history due to their rather intuitive justification based on entropy's properties as an information measure. Recently, however, Shore and Johnson have demonstrated (Shore, 1980) that these principles are correct general methods of inference when given information in terms of expected values. Their results rest on four consistency axioms which are used

to demonstrate maximizing any other function, but entropy will lead to logical inconsistencies unless that function and entropy have identical maxima.

### Definitions and General Problem Statement

Given the historical outline, this section will describe the general setting where the maximum entropy and minimum cross-entropy principles can be applied and define the notation that will be used throughout the dissertation. The main interest in this work is approximating continuous bivariate density functions and making logical inferences based on this approximation. Because the cluttered aerial scene problem is driving this review, all  $n$ -dimensional results will only be presented for bivariate density functions.

The theory for approximating discrete probability density functions using the principle of maximum entropy is well-known and has found a great many applications. In this problem formulation, the underlying system has  $n$  possible states  $x_i$  and they occur with unknown probabilities  $q(x_i)$ . The system is observed with the observations taking the form  $\sum_i q(x_i) f_k(x_i) = m_k$  or as the expected value of a set of "information functions"  $\{f_k\}$ . The problem then is to choose a distribution  $e(x_i)$  that is in some sense the best estimate of  $q(x_i)$  given the expected value measurements. In general, there remains an infinite set of distributions that are not ruled out by the expected value measurements that now serve as constraints on any approximating distribution.

The maximum entropy principle, however, provides a unique approximation density  $e(x_i)$  by selecting from the infinite set of densities that satisfy the constraints the one density with the largest entropy defined as  $-\sum_i e(x_i) \log[e(x_i)]$ .

The principle of minimum cross-entropy is a generalization of the maximum entropy principle that applies in cases when a prior distribution  $p(x_i)$  that estimates  $q(x_i)$  is known in addition to the measurement constraints. The principle states that: of the infinite set of distributions  $e(x_i)$  that satisfy the constraints, choose the one with the least cross-entropy  $\sum_i e(x_i) \log[e(x_i)/p(x_i)]$ . The connection between the two principles occurs when the prior is a uniform density and in this case minimizing cross-entropy is equivalent to maximizing entropy. The concept of cross-entropy also generalizes correctly for continuous probability densities unlike the concept of maximum entropy, where only a differential entropy is defined in the continuous case and that is not even invariant under coordinate transformations (McEliece, 1977:38).

In the case of continuous bivariate probability densities the principle of minimum cross-entropy provides a general method of inference about an unknown density  $q(x,y)$  when there exists a prior estimate of  $q(x,y)$  and new information about the unknown density in the form of expected values of the information functions. The principle states that: of all the densities that satisfy the expected value constraints,

choose as the approximating density the posterior  $e(x,y)$  with the least cross-entropy

$$H(e,p) = \iint_D e(x,y) \log \left[ \frac{e(x,y)}{p(x,y)} \right] dx dy$$

where  $p(x,y)$  is a prior estimate of  $q(x,y)$ . Jaynes has also shown (Jaynes, 1968) that generalizing entropy maximization to continuous densities leads to the above cross-entropy functional with  $p(x,y)$  being called an "invariant measure" instead of a prior density. When using the entropy maximization principle, there is an implicit assumption of uniform priors when viewed from the broader cross-entropy perspective. The failure of maximum entropy to generalize as might be expected is also explained by this viewpoint since a uniform prior in one coordinate system may not be uniform in another coordinate system (Shore, 1980).

#### The Consistency Axioms

Shore and Johnson (Shore, 1980) have proven that given a prior density and new information in the form of constraints on expected values, there is only one posterior density satisfying these constraints that can also be chosen in a manner that satisfies a set of logical consistency axioms. In addition, this unique posterior density can be obtained by minimizing the cross-entropy functional. The four consistency axioms are informally defined as follows:

1. Uniqueness: The result should be unique.
2. Invariance: New information can be accounted for in any coordinate system.
3. System Independence: Independent information about independent systems can be accounted for separately in terms of different densities or together in terms of a joint density.
4. Subset Independence: Information about an independent subset of system states can be accounted for in terms of a separate conditional density or in terms of the full system density.

All four of these axioms are based on a single fundamental principle: If a problem can be solved in more than one way, the results should be consistent (Shore, 1980). The axioms are the desired properties of an inference procedure rather than the desired properties of an information measure. Using only a general functional  $J(e,p)$  to select the posterior density  $e(x,y)$  in the inference procedure and starting with the axioms of subset independence and invariance, Shore and Johnson were able to show that the first consequence of their axioms was to restrict  $J(e,p)$  to functionals that are equivalent to the form

$$J(e,p) = \iint_D f[e(x,y), p(x,y)] dx dy$$

for some function  $f$  of two variables. This functional form is called the "sum form" and in work previous to Shore and Johnson's development, the sum form was assumed rather than derived (Johnson, 1979). Then having established this functional form and using the general axiom of invariance, they show that  $J$  is further restricted to functionals that are equivalent to the form

$$J(e,p) = \iint_D e(x,y) h\left[\frac{e(x,y)}{p(x,y)}\right] dx dy$$

where  $h$  is some function of a single variable. Using all four axioms, Shore and Johnson are finally able to show that  $J$  must be equivalent to the functional

$$J(e,p) = \iint_D e(x,y) \log\left[\frac{e(x,y)}{p(x,y)}\right] dx dy$$

or  $J(e,p)$  must be equivalent to cross-entropy. Since it is possible that no functional satisfies the consistency axioms, their final step is to show that the cross-entropy functional  $H(e,p)$  satisfies all four axioms. The Shore and Johnson result has immediate application to approximating the normalized irradiance distribution  $i(x,y)$  since it provides a logically consistent method of approximating the light density based on measurements in the form of expected values. The procedure to follow then requires a prior estimate of the light density, expected value information about

the true density  $i(x,y)$  and the functional  $H(e,p)$  to measure how much the prior density differs from the posterior density. The principle of minimum cross-entropy is then the correct method of incorporating all the given information and producing a logically consistent posterior density  $e(x,y)$  that approximates the unknown true light density  $i(x,y)$ .

### Properties of Cross-Entropy Minimization

The basic properties of cross-entropy minimization are fundamental to the problem of detecting objects in a cluttered aerial scene using the posterior density  $e(x,y)$  as an optimum light density approximation. Because of their importance in developing a target detection algorithm and for completeness, I will outline the well-known properties of cross-entropy minimization and the notational system developed by Johnson and Shore (Johnson, 1980). Many results dealing with cross-entropy minimization can be efficiently stated in terms of an abstract information operator  $*$  which takes the two known arguments of a prior density and new expected value information to yield a posterior density. Using this operator notation, the posterior  $e$  is given by  $e = p * I$  where  $I$  stands for the known constraints on the expected values.

The problem will be stated more formally in this section to allow concise definitions of minimum cross-entropy properties. Again in this outline, because of the thrust in this dissertation of approximating a bivariate density function, all results will be presented only for the two-dimensional

case. The formal problem statement defines a point  $(x,y)$  in the  $x$ - $y$  plane as a system state with  $D$  the region in the plane where all states are defined. Then  $S$  is the set of all probability densities  $S(x,y)$  on  $D$  such that  $S(x,y) \geq 0$  for  $(x,y) \in D$  and

$$\iint_D S(x,y) dx dy = 1$$

New information takes the form of linear equality constraints or

$$\iint_D q(x,y) f_k(x,y) dx dy = m_k$$

where  $q(x,y)$  is the unknown true system density with  $q(x,y) \in S$  and  $f_k(x,y)$  are known information functions with known expected values. The probability densities that satisfy these constraints always comprise a convex subset  $Z$  of  $S$  (Johnson, 1979). The set  $Z$  is then termed a constraint set and in general, a given convex region  $Z$  of  $S$  may be defined by more than one set of information functions. The fact that the constraints form a convex subset of  $S$  insures the convergence of computational methods attempting to find the minimum cross-entropy posterior density. The expected value constraints and the resulting convex set  $Z$  form the term  $I$  used in the abstract operator notation.

The second argument for the information operator  $*$  is



a prior density  $p(x,y)$ . The density  $p(x,y)$  contained in  $S$  is used to define an estimate of  $q(x,y)$  which can be obtained by any means prior to learning the average value information  $I$ . The prior density is required to be strictly positive over  $D$ :

$$p[(x,y) \in D] > 0$$

In making this restriction, it is assumed that  $D$  is the set of states that is possible according to the prior information. The restriction does not significantly restrict results, but does avoid the technical problems that would result from division by  $p(x,y)$  equal to zero. In a more general setting,  $D$  would be a measurable space and  $p$  and  $e$  would be replaced by prior and posterior probability measures. By defining probability densities, it is implicitly assumed there is some underlying measure with respect to which the other measures are absolutely continuous (Kullback, 1959:4). Such a measure will exist when no event with zero prior probability can have a positive posterior probability and which is demanded by the strictly positive assumption for  $p(x,y)$  (Giuasu, 1977).

Given the two arguments for the information operator  $*$  (the prior  $p(x,y)$  and new information  $I$ ), the posterior density  $e(x,y) \in Z$  that results from taking  $I$  into account is selected by minimizing the cross-entropy  $H(s,p)$  in the

constraint set  $Z$ :

$$H(e,p) = \min_{s \in Z} H(s,p)$$

Using this problem statement and the cross-entropy minimization procedure, the following properties apply to cross-entropy minimization:

Property 1: (Uniqueness) The posterior  $e = p*I$  is unique.

The uniqueness property insures that the solution of a given cross-entropy minimization problem for the posterior density  $e(x,y)$  is unique. The minimization of the functional  $H(e,p)$  allows this unique density to be identified.

Property 2: (Prior Omnipotence) The posterior satisfies  $e = p*I = p$  if and only if the prior satisfies  $p \in Z$ .

The prior omnipotence property shows that when the new information  $I$  agrees with the assumed prior density the prior and posterior are equal. When cross-entropy minimization is viewed as an inference procedure, it makes sense that the posterior density  $e(x,y)$  should be unchanged from the prior, if the new information does not contradict the prior density  $p(x,y)$  in any way.

Property 3: (Idempotence)  $(p*I)*I = p*I$

Idempotence insures that taking the same information into account twice has the same effect as taking it into account once.

Property 4: (Information Intersection) Let  $I$  be the information  $I_1 = (q \in Z_1)$  where this notation denotes that  $q$  is a member of the constraint set  $Z_1 \subseteq S$  created by the constraints  $I_1$  and  $I_2$  the information  $I_2 = (q \in Z_2)$ , for overlapping constraint sets  $Z_1, Z_2 \subseteq S$ . If  $(p * I_1) \in Z_2$  holds, then

$$p * I_1 = (p * I_1) * (I_1 \cap I_2) = (p * I_1) * I_2 = p * (I_1 \cap I_2)$$

holds.

The information intersection property is similar to the prior omnipotence property. The result shows that when  $I_1$  is taken into account, if the resulting posterior density  $p * I_1$  already satisfies the constraints imposed by the additional information  $I_2$ , then taking  $I_2$  into account in various ways has no effect on the posterior density.

Property 5: (Invariance) Let  $T$  be a coordinate transformation from  $(x, y) \in D$  to  $(u, v) \in R$  with  $(Te)(u, v) = J^{-1} e(x, y)$ , where  $J$  is the Jacobian  $J = \partial(u, v) / \partial(x, y)$ . Let  $TS$  be the set of densities  $Te$  corresponding to densities  $e \in S$ . Let  $(TZ) \subseteq (TS)$  correspond to  $Z \subseteq S$ . Then

$$(Tp)*(TI) = T(p*I)$$

and

$$H[T(p*I), Tp] = H(p*I, p)$$

hold, where

$$TI = [(Tq) \in (TZ)]$$

or  $Tq$  is a member of the constraint set  $TZ \subseteq TS$  created by the constraints  $TI$ .

The invariance property states that the same answer is obtained when an inference problem is solved in two different coordinate systems, in that the posterior densities in the two systems are related by the coordinate transformation. Also, the cross-entropy between the posteriors and the priors has the same value in both coordinate systems.

Property 6: (System Independence) Let there be two systems, with sets  $D_1$  and  $D_2$  of states and probability densities of states  $e_1 \in S_1$  and  $e_2 \in S_2$ . Let  $p_1 \in S_1$  and  $p_2 \in S_2$  be prior densities. With  $I_1 = (q_1 \in Z_1)$  and  $I_2 = (q_2 \in Z_2)$  new information about the two systems, where  $Z_1 \subseteq S_1$  and  $Z_2 \subseteq S_2$ . Then

$$(p_1 p_2)*(I_1 \cap I_2) = (p_1 * I_1)(p_2 * I_2)$$

and

$$H(e_1 e_2, p_1 p_2) = H(e_1, p_1) + H(e_2, p_2)$$

hold where  $e_1 = p_1 * I_1$  and  $e_2 = p_2 * I_2$ .

The system independence property shows that it does not matter whether independent information about two systems is accounted for separately or together in terms of a joint density. Whether or not the two systems are in fact independent is irrelevant since the property applies as long as there are independent priors and independent new information.

Property 7: (Triangle Relations) For any  $r(x,y) \in Z$

$$H(r,p) \geq H(r,e) + H(e,p)$$

where  $e = p * I$ . When  $I$  is determined by a finite set of equality constraints only, equality holds.

The triangle equality is important for all applications in which cross-entropy minimization is used for purposes of classification on pattern recognition.

Property 8: (Posterior Convergence) The relationship

$$H(q, p * I) \leq H(q, p)$$

holds with equality, if and only if  $p * I = p$ .

The posterior convergence property states that the posterior density  $e(x,y)$  is always closer to the scene density  $q(x,y)$  in the cross-entropy sense than is the prior density  $p(x,y)$ .

Property 9: (Piecemeal Information) Let the system have a probability density  $q \in S$ , and let there be information  $I_1 = (q \in Z_1)$  and  $I_2 = (q \in Z_2)$ , where  $Z_1, Z_2 \subseteq S$  are constraint sets with non-empty intersection. Given that  $Z_1$  is determined by a set of equality constraints only, then

$$(p * I_1) * (I_1 \cap I_2) = p * (I_1 \cap I_2)$$

and

$$H(e, p) = H(e, e_1) + H(e_1, p)$$

hold where  $e = p * (I_1 \cap I_2)$  and  $e_1 = p * I_1$ .

The piecemeal information property is also important because of its application in classification and pattern recognition. In general, this result is important in any application where the constraint information arrives piecemeal and states that intermediate posterior densities can be used as priors in computing final posterior densities without affecting the results.

There are additional cross-entropy minimization proper-

ties of general interest not covered in this listing. The additional properties will be developed and discussed in the next chapter. Chapter III will develop a minimum cross-entropy posterior density approximation and a target detection algorithm based on the approximation and properties of minimum cross-entropy densities.

### Chapter III. Detection Algorithm Development

#### Introduction

The theoretical development of an algorithm for detecting complex man-made objects in cluttered aerial scenes will be presented in this chapter. The introduction to this dissertation outlined the general framework for the object detection problem and showed how scene frames are partitioned in  $K^2$  information cells. The detection algorithm developed in this chapter is then sequentially applied to each information cell in a frame resulting in all cells being classified as containing targets or only clutter. To develop the detection algorithm, the irradiance distribution function for the  $i$ th cell, in the  $j$ th frame will be denoted  $Q_{ij}(x,y)$ . The normalized irradiance distribution function is denoted  $q_{ij}(x,y)$  and has all the properties of a bivariate probability density function. Following the notation of previous chapters,  $e_{ij}(x,y)$  is the minimum cross-entropy approximation to the  $i$ th cell and  $j$ th frame normalized irradiance distribution function  $q_{ij}(x,y)$ . The computation of the approximation  $e_{ij}(x,y)$  requires a prior  $i$ th cell and  $j$ th frame density  $p_{ij}(x,y)$  and new expected value information  $I_{ij}$ . Throughout the remaining sections of this dissertation it is assumed we are working with the  $i$ th information cell, in the  $j$ th frame of an aerial scene and the explicit reference



to the cell and frame number will be dropped unless it is required for clarity.

### Solution of the Constraint Equations

To classify cell density functions  $q_{ij}(x,y)$  as containing targets or only clutter will require an explicit procedure for obtaining the minimum cross-entropy density  $e_{ij}(x,y)$  based on a set of information function expected value relations. The information functions  $f_k(x,y)$  used in the minimum cross-entropy inference procedure are critical components of the detection algorithm and will be explored fully in the next chapter. The expression for the minimum cross-entropy posterior density can be found given that the number and forms of the information functions are specified and their expected values have been computed over the information cell or symbolically, given  $f_k(x,y)$  and  $m_k$ ;  $k = 0,1,2\dots t$  are known. The minimum cross-entropy posterior approximation of  $q(x,y)$  will then be the continuous density  $e(x,y)$  defined on the region  $-C \leq x,y \leq C$  that has a prior representation  $p(x,y)$  and will satisfy the new expected value information I. The mathematical statement of the problem is to find  $e(x,y)$  subject to the constraints:

$$\min\{H(e,p)\} = \max\left\{-\int_{-C}^C \int_{-C}^C e(x,y) \ln \left[ \frac{e(x,y)}{p(x,y)} \right] dx dy \right\}$$

subject to

$$\int_{-c}^c \int_{-c}^c e(x,y) dx dy = 1$$

and

$$\int_{-c}^c \int_{-c}^c f_k(x,y) e(x,y) dx dy = m_k$$

$$k = 1, 2, \dots, t$$

The information functions  $f_k(x,y)$ ,  $k = 1, 2, \dots, t$  are continuous and bounded on the region  $-C \leq x, y \leq C$ . The problem stated above is a constrained minimum problem and can be solved using the Lagrange method of undetermined coefficients. The Lagrangian,  $L[e(x,y), \Lambda]$  is then formed as follows (Luenberger, 1969:213):

$$L[e(x,y), \Lambda] = -H(e, p) - \lambda_0 \left\{ \int_{-c}^c \int_{-c}^c e(x,y) dx dy - 1 \right\}$$

$$- \sum_{j=1}^t \lambda_j \left\{ \int_{-c}^c \int_{-c}^c f_j(x,y) e(x,y) dx dy - m_j \right\}$$

Using the expression for cross-entropy, the Lagrangian can be expressed as:

$$L[e(x,y), \Lambda] = \int_{-c}^c \int_{-c}^c e(x,y) \left\{ \ln \left[ \frac{p(x,y)}{e(x,y)} \right] - \lambda_0 - \sum_{j=1}^t \lambda_j f_j(x,y) \right\} dx dy$$

$$+ \lambda_0 + \sum_{j=1}^t \lambda_j m_j$$

The Lagrangian can also be written in the form:

$$L[e(x,y), \Lambda] = \iint_{-c}^c e(x,y) \left\{ \ln \left[ \frac{p(x,y)}{e(x,y)} \cdot \exp \left[ -\lambda_0 - \sum_{j=1}^t \lambda_j f_j(x,y) \right] \right] \right\} dx dy \\ + \lambda_0 + \sum_{j=1}^t \lambda_j m_j$$

Now using the fact that for  $Z > 0$  the natural logarithm is bounded by

$$\ln(Z) < Z - 1 \quad \text{if } Z \neq 1$$

and

$$\ln(Z) = Z - 1 \quad \text{if } Z = 1$$

provides a method of bounding the Lagrangian. Using this property of natural logarithms provides the relationship:

$$L[e(x,y), \Lambda] \leq \iint_{-c}^c e(x,y) \left\{ \frac{p(x,y)}{e(x,y)} \cdot \exp \left[ -\lambda_0 - \sum_{j=1}^t \lambda_j f_j(x,y) \right] - 1 \right\} dx dy \\ + \lambda_0 + \sum_{j=1}^t \lambda_j m_j$$

The goal of this procedure is to maximize the Lagrangian  $L[e(x,y), \Lambda]$  and therefore  $e(x,y)$  must be selected to provide equality in the last expression. Again, using the property of natural logarithms equality occurs, if and only if

$$e(x,y) = p(x,y) \exp \left[ -\lambda_0 - \sum_{j=1}^t \lambda_j f_j(x,y) \right]$$

the point where  $Z = 1$ . The preceding result is well-known (Johnson, 1980:27), however, the derivation given here is unique to this dissertation. The derivation of the minimum cross-entropy density seen here is an extension of Miller's work in maximum entropy and univariate densities (Miller, 1980:29).

The expression above for  $e(x,y)$  provides the required form of the minimum cross-entropy density that approximates the unknown true density  $q(x,y)$ . Given a specific set of expected values  $(m | k = 0, 1, \dots, t)$ , we solve the  $t + 1$  constraint equations for  $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_t)^T$  to then completely determine  $e(x,y)$ . The method of solving the given set of nonlinear constraint equations for the lambda vector will be presented in Chapter V.

From property one of Chapter II, we know the minimum cross-entropy posterior density  $e(x,y)$  is unique. In terms of the abstract information operator  $*$  a solution to the cross-entropy minimization problem, if one exists, is unique provided only that  $H(e,p)$  is not identically infinite as  $e(x,y)$  ranges over the constraint set  $Z$ . A condition that guarantees the existence of a solution is that in addition to containing a density  $e(x,y)$  with finite cross-entropy, the constraint set  $Z$  is closed (Johnson, 1980:5). For  $Z$  to be closed, it suffices in turn that the constraint functions  $f_k(x,y)$  are bounded. Conversely, given values of  $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_t)^T$  such that all constraints are satisfied,

then the solution exists and is given by the above minimum cross-entropy expression for  $e(x,y)$ . Conditions for the general existence of solutions to the constrained minimization problem are also discussed by Csiszar (Csiszar, 1975). For this work with normalized irradiance distributions using a finite set of bounded information functions  $f_k(x,y)$  and only equality constraints, the solution to the constrained minimization problem will always exist and have the unique form for  $e(x,y)$  derived in this section as the minimum cross-entropy density.

#### Solution Characteristics

In general cross-entropy  $H(e,p)$  measures how much  $e(x,y)$  differs from the prior  $p(x,y)$ . The cross-entropy at the minimum can be expressed in terms of the Lagrange multipliers and the expected values of the information functions. Starting with the expression for the minimum cross-entropy density or

$$e(x,y) = p(x,y) \exp \left[ -\lambda_0 - \sum_{i=1}^t \lambda_i f_i(x,y) \right]$$

and rearranging gives the expression

$$\ln \left[ \frac{e(x,y)}{p(x,y)} \right] = \left[ -\lambda_0 - \sum_{i=1}^t \lambda_i f_i(x,y) \right]$$

Now multiplying by  $e(x,y)$  and integrating over the information cell gives the expression

$$\iint e(x,y) \ln \left[ \frac{e(x,y)}{p(x,y)} \right] dx dy = - \lambda_0 \iint e(x,y) dx dy$$

$$- \sum_{j=1}^t \lambda_j \iint f_j(x,y) e(x,y) dx dy$$

Therefore, the cross-entropy  $H(e,p)$  at the minimum point is given by

$$H(e,p) = -\lambda_0 - \sum_{j=1}^t \lambda_j m_j$$

Kullback has also shown that cross-entropy in general satisfies the relationship

$$H(e,p) \geq 0$$

and with equality only if  $p(x,y) = e(x,y)$  almost everywhere (Kullback, 1959). Informally,  $H(e,p)$  is a measure of the information divergence between the density function  $e(x,y)$  and a prior density function  $p(x,y)$ . Then using  $H(e,p)$  as an information divergence measure and since  $e = p^*I$  minimizes  $H(e,p)$ , the posterior approximation for  $q(x,y)$  is as close as possible in an information-measure sense to the prior density while at the same time satisfying the new information constraints  $I$  taken from the unknown cell density  $q(x,y)$ .

#### Further Minimum Cross-Entropy Properties

The properties presented in this section highlight

cross-entropy's ability to measure how much a posterior density differs from the assumed prior density. Even though cross-entropy does not have all the properties of a metric,  $H(e,p)$  does have other properties that make it ideal for use in a target detection algorithm. These properties are presented in this section and then used in the next section to develop the minimum cross-entropy detection rule.

Triangle Equality: Let  $I$  be the constraints given by

$$\iint_{\text{cell}} f_k(x,y)q(x,y)dxdy = m_k$$

$$k = 1,2,\dots,t$$

and let  $p(x,y)$  be any prior probability density. Then

$$H(q,p) = H(q,p*I) + H(p*I,p)$$

The minimum cross-entropy posterior estimate of  $q(x,y)$  is both logically consistent (four consistency axioms) and closer to  $q(x,y)$  as measured by cross-entropy than the prior density  $p(x,y)$ . Also, the difference  $H(q,p) - H(q,e)$  is exactly the cross-entropy  $H(e,p)$  between the posterior and the prior. Therefore,  $H(e,p)$  can be interpreted as the amount of information provided by the constraints  $I$  that is not inherent in  $p(x,y)$ . The posterior accessibility property also shows that the difference  $H(q,p) - H(q,e)$  will

equal zero when the correct expected value constraints are provided.

Posterior Accessibility: For any density  $d(x,y)$  there exists constraints  $I_d$  such that  $d = p * I_d$  for any prior density  $p(x,y)$ .

This property due to Csiszar (Csiszar, 1975) shows that  $H(d,p)$  is in general the amount of information needed to determine  $d(x,y)$  when given the prior  $p(x,y)$ . The result also shows that the cross-entropy  $H(d,p)$  measures the error introduced by using  $p(x,y)$  instead of the true density  $d(x,y)$ . Used as an error measure, the posterior accessibility property will allow the template to scene cross-entropy  $H(q,t)$  to provide a "metric" for measuring the detection rule's sensitivity to variations in the performance determining parameters presented in Chapter V. The next property also shows that the minimum cross-entropy template provides the minimum error possible when the template is restricted to an exponential form.

Expected Value Matching: Let  $I$  be the constraints

$$\iint_{\text{cell}} f_k(x,y) q(x,y) dx dy = m_k$$

$$k = 1, 2, \dots, t$$

for a fixed set of information functions  $f_k(x,y)$  and let



$e = p \cdot I$  be the result of taking this information  $I$  into account. Then for an arbitrary fixed density  $d(x,y)$  the cross-entropy  $H(d,e) = H(d,p \cdot I)$  has its minimum value when the constraints satisfy

$$m_k = m_k^{(d)} = \iint_{\text{cell}} d(x,y) f_k(x,y) dx dy$$

$$k = 1, 2, \dots, t$$

This result is due to Johnson and Shore (Johnson, 1980) and is a generalization of a property of orthogonal polynomials that in the case of speech analysis is called the "correlation matching property" (Markel, 1976). Using this result insures that when a minimum cross-entropy density  $e(x,y)$  of the general form

$$e(x,y) = p(x,y) \exp \left[ -\lambda_0 - \sum_{j=1}^t \lambda_j f_j(x,y) \right]$$

then  $H(d,e)$  is smallest when the expectations of  $e(x,y)$  match those of the arbitrary density  $d(x,y)$ . Therefore, in general it follows that  $e = p \cdot I$  is not only the density that minimizes the prior to posterior cross-entropy  $H(e,p)$ , but it is also the density of the general form shown above that minimizes the posterior to scene cross-entropy  $H(q,e)$  since  $d(x,y)$  was an arbitrary density (Shore, 1980). Hence,  $e(x,y)$  is not only closer to  $q(x,y)$  than is  $p(x,y)$ , but

it is the closest possible density of the exponential form given above for  $e(x,y)$ .

The final property of cross-entropy minimization required to develop a target in clutter detection algorithm is the posterior adaptation property presented by Johnson and Shore (Johnson, 1980).

Posterior Adaptation: Let  $I_1$  and  $I_2$  stand respectively for the information constraints

$$\iint_{\text{cell}} f_j(x,y) q_1(x,y) dx dy = m_j^{(1)}$$

and

$$\iint_{\text{cell}} f_j(x,y) q_2(x,y) dx dy = m_j^{(2)}$$

which involve the same set of information functions  $f_j(x,y)$  where  $j = 1, 2, \dots, t$ . Then

$$(p * I_1) * I_2 = p * I_2$$

and

$$H(e_2, p) = H(e_2, e_1) + H(e_1, p) + \sum_{j=1}^t \lambda_j^{(1)} \left[ m_j^{(1)} - m_j^{(2)} \right]$$

hold where  $e_1 = p * I_1$ ,  $e_2 = p * I_2$  and  $\lambda_j^{(1)}$  are the Lagrange multipliers associated with  $e_1 = p * I_1$ .

The application of this property views  $q_1(x,y)$  and  $q_2(x,y)$  as the unknown system or scene probability densities at two different points in time. Then  $q_1(x,y)$  is used as a prior estimate or template for  $q_2(x,y)$ . The posterior adaptation property shows that when  $I_2$  is determined by expectations of the same information functions that also produced  $I_1$ , the results of producing a posterior  $e_1(x,y)$  using  $I_1$  are completely wiped out by subsequently producing a posterior  $e_2(x,y)$  using  $I_2$ . The posterior adaptation property is shown graphically in Figure 3.1.

#### The Detection Algorithm

Using a constant set of information functions  $[f_j(x,y) | j = 0,1,2...t]$  (see Chapter IV) and a uniform prior density  $p(x,y)$ , the posterior adaptation property serves as a starting point for the detection algorithm. The information  $I_1^{(k)}$  is obtained from a set of predefined template scenes  $q^{(k)}(x,y)$  where  $k = 1,2...2Q$ . These template scenes model the target of interest and various possible clutter configurations to provide the detection rule with  $Q$  target versus clutter alternatives. With this information  $I_1^{(k)}$  a set of minimum cross-entropy (maximum entropy) template densities  $t^{(k)}(x,y)$  where  $k = 1,2...2Q$  can be defined as  $t^{(k)} = p * I_1^{(k)}$  corresponding to the  $e_1(x,y)$  density in the posterior adaptation result.

In a more general setting, when there are  $N$  targets of interest the minimum cross-entropy template densities will

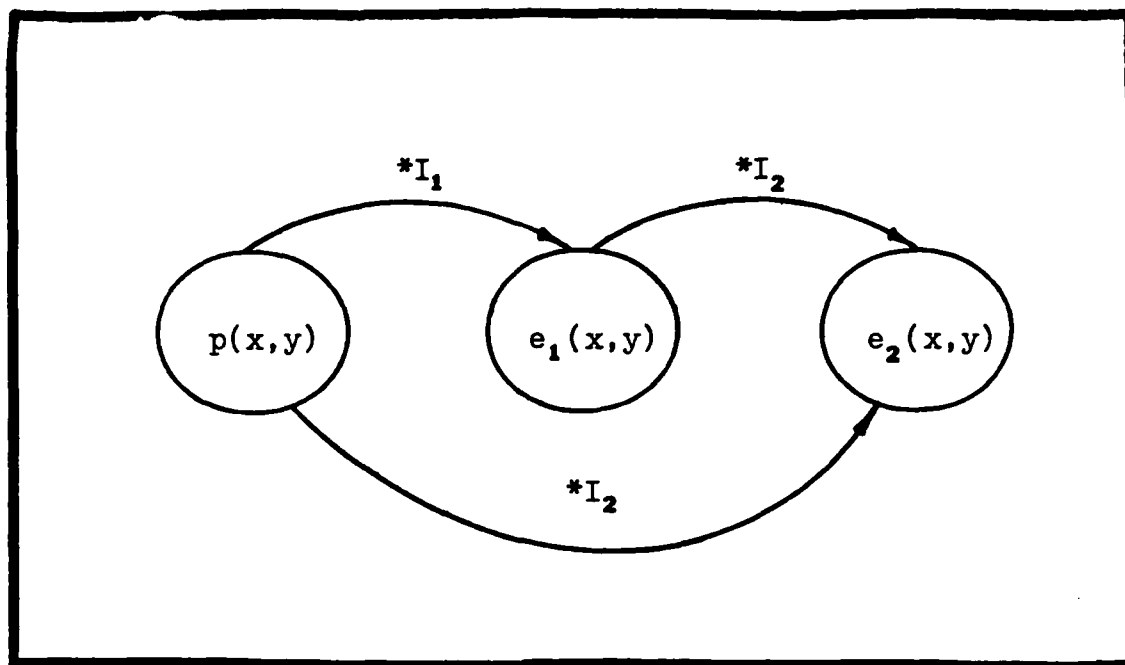


Fig. 3.1. Posterior Adaptation

be denoted

$$t^{(lk)}(x,y) \text{ where } l = 0,1,2\dots N$$

$$\text{and } k = 1,2\dots Q$$

This notation refers to the  $N$  targets of interest as discussed in Chapter I where zero is reserved for the "only clutter" template and there are  $Q$  possible clutter backgrounds for each of the total  $N + 1$  target classes. The resulting  $(N + 1)Q$  template densities are obtained from a minimum cross-entropy procedure as

$$t^{(lk)} = p * I^{(lk)}$$

where  $p(x,y)$  is the uniform prior density and  $I^{(lk)}$  is the expected value information obtained from the template scenes as

$$\int\int_{\text{cell}} f_j(x,y) q^{(lk)}(x,y) dx dy = m_j^{(lk)}$$

where

$$j = 1, 2 \dots t$$

$$k = 1, 2 \dots Q$$

and

$$l = 0, 1 \dots N$$

The production of the  $Q(N + 1)$  template scenes  $q^{(lk)}(x,y)$  is presented in Chapter VI, however, the basic principle uses a master target template to represent each of the  $N$  target classes. Each master target template is then superimposed on the  $Q$  different clutter backgrounds to offer  $Q$  target and clutter configurations to the detection algorithm for each target class. Appendix A shows eighteen three-dimensional template scenes used to test the detection rule, nine of which represent a tank in clutter and nine of which represent only clutter. This set of template scenes where  $N$  equals 1 and  $Q$  equals 9 provided the expected value information used to produce the template densities shown in

Appendix B. In the general setting and using the resulting template densities  $t^{(k)}(x,y)$  every information cell analyzed by the detection rule will be classified as only clutter or as containing one of the  $N$  targets of interest. The classification is based on the ability of cross-entropy  $H(q,t^{(k)})$  to measure how much the cell density  $q(x,y)$  differs from the template densities  $t^{(k)}(x,y)$ .

To develop the actual classification rule several other minimum cross-entropy properties must also be used. Using the same set of information functions  $[f_j(x,y) | j = 0,1,2...t]$  used to construct the minimum cross-entropy template densities  $t^{(k)}(x,y)$ , measurements of the information functions expected values are taken from the scene density  $q_{ip}(x,y)$  of the  $i$ th cell and the  $p$ th frame of the aerial scene. These measurements form a set of constraints  $I$  on the posterior density and are obtained as

$$\iint_{\text{cell}} f_j(x,y) q(x,y) dx dy = m_j$$

where

$$j = 1,2...t$$

to form a measurement vector  $M$ . Using this constraint information coupled with the prior density  $p(x,y)$  will allow a minimum cross-entropy posterior density to be produced as

$e = p * I$ . Figure 3.2 shows all the densities used in this development of the detection algorithm and how these densities evolve as new expected value information is used in the minimum cross-entropy procedure.

The cell information  $I$  can also be applied to the  $(N + 1)Q$  minimum cross-entropy template densities  $t^{(lk)}(x,y)$ . Abstractly, this procedure is forming a new set of  $(N + 1)Q$  adapted template densities using the new expected value information  $I$  and the predefined template densities  $t^{(lk)}(x,y)$  as priors. Using the operator notation adapted densities are constructed as

$$d^{(lk)} = t^{(lk)} * I$$

where

$$l = 0, 1, 2 \dots N$$

$$k = 1, 2 \dots Q$$

Figure 3.2 provides a complete summary of the detection algorithm notation and minimum cross-entropy densities being generated. The scene density  $q(x,y)$  represents a general information cell that must be classified by the algorithm.

Now the triangle equality can be applied to show that

$$H(q, t^{(lk)}) = H(q, d^{(lk)}) + H(d^{(lk)}, t^{(lk)})$$

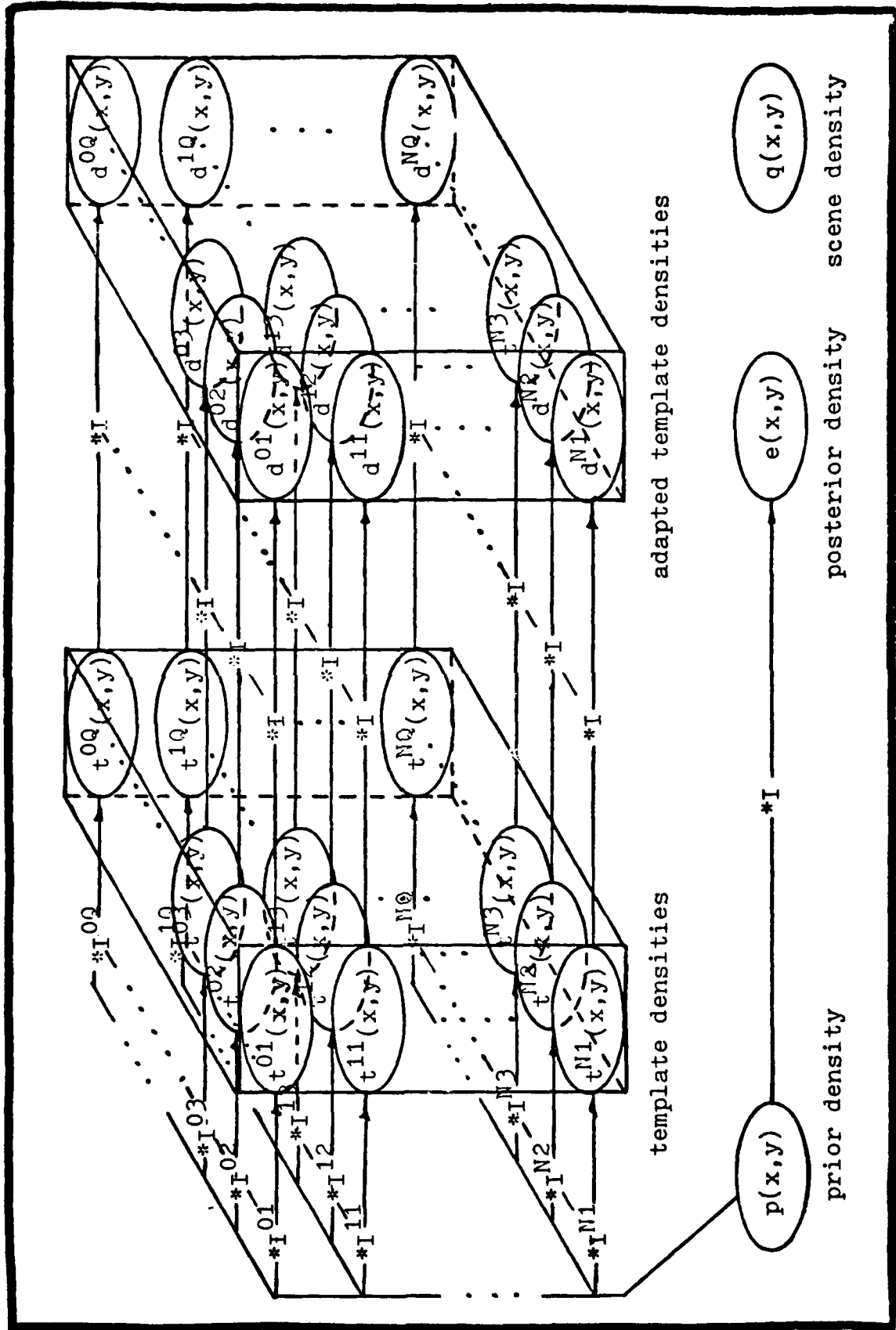


Fig. 3.2. Density Function Evolution



where  $l = 0, 1, 2, \dots, N$  and  $k = 1, 2, \dots, Q$ . Also the adapted template densities were obtained as

$$d^{(lk)} = t^{(lk)} * I$$

or

$$d^{(lk)} = (p * I^{(lk)}) * I$$

Using the posterior adaptation property results in the expression

$$d^{(lk)} = p * I = e$$

where again  $l = 0, 1, 2, \dots, N$  and  $k = 1, 2, \dots, Q$ . This result shows that all  $(N + 1)Q$  adapted template densities shown in Figure 3.2 are equal to the single posterior density  $e(x, y)$ . Returning to triangle equality with this result gives

$$H(q, t^{(lk)}) = H(q, e) + H(e, t^{(lk)})$$

with  $l = 0, 1, 2, \dots, N$  and  $k = 1, 2, \dots, Q$ .

The cross-entropy  $H(q, t^{(lk)})$  is the amount of information needed to determine the true information cell density  $q(x, y)$  given the predefined template density  $t^{(lk)}(x, y)$  or it is a measure of how much  $q(x, y)$  differs from the template

density. The triangle equality given above shows that the total "distance" between  $q(x,y)$  and the template density  $t^{(k)}(x,y)$  or  $H(q,t^{(k)})$  is the sum of two components. The expected value matching property has shown that the first term  $H(q,e)$  already represents a minimum "distance" between  $q(x,y)$  and the best posterior estimate of  $q(x,y)$  with the required exponential form. The second variable term  $H(e,t^{(k)})$  is the "distance" from the template density  $t^{(k)}(x,y)$  to the minimum cross-entropy posterior density  $e(x,y)$ .

The strategy for a detection algorithm is now to use the expected value matching property of the minimum cross-entropy procedure. Also, since  $H(q,e)$  has previously been shown to have its minimum possible value, the detection rule must select the template density  $t^{(\alpha\beta)}(x,y)$  from the set of  $(N + 1)Q$  total templates because it is the density that minimizes the cross-entropy  $H(e,t^{(\alpha\beta)})$ . The triangle equality for cross-entropy therefore results in a detection rule based on the "distance" between minimum cross-entropy template densities  $t^{(k)}(x,y)$  and the minimum cross-entropy scene density  $e(x,y)$ . The rule requires that we find  $\alpha\beta$  such that

$$H(e,t^{(\alpha\beta)}) < H(e,t^{(k)})$$

as  $\alpha$  varies from 0 to  $N$  and  $k$  varies from 1 to  $Q$ . The detection rule that results is equivalent to the classi-

fication rule given by Kullback (Kullback, 1959:83). Using this detection rule results in finding the template density  $t^{(a\beta)}(x,y)$  that differs the least from the true information cell density  $q(x,y)$ . The numerical value of  $a$  will then indicate if the  $i$ th information cell in the  $p$ th frame contains only clutter or one of the  $N$  objects of interest.

Now to implement the detection rule numerically a modification of a result provided by Gray and Shore (Gray, 1980) for a speech coding technique will be used. The second result from the posterior adaptation property stated that

$$H(e_2, p) = H(e_2, e_1) + H(e_1, p) + \sum_{j=1}^I \lambda_j^{(1)} (m_j^{(1)} - m_j^{(2)})$$

Rearranging and changing to the target density notation gives

$$H(e, t^{(k)}) = H(e, p) - H(t^{(k)}, p) - \sum_{j=1}^I \lambda_j^{(k)} (m_j^{(k)} - m_j)$$

where  $\lambda_j^{(k)}$  are the Lagrange multipliers used in the pre-defined template density  $t^{(k)}(x,y)$ . The template densities are also obtained through a minimum cross-entropy procedure and therefore

$$H(t^{(k)}, p) = -\lambda_0^{(k)} - \sum_{j=1}^I \lambda_j^{(k)} m_j^{(k)}$$

is the cross-entropy at the minimum. Then substitution of this expression into the  $H(e, t^{(k)})$  expression gives

$$H(e, t^{(rs)}) = H(e, p) + \lambda_0^{(rs)} + \sum_{i=1}^t \lambda_i^{(rs)} m_i$$

The term  $H(e, p)$  is a constant for all template densities and will not enter into the decision rule. The detection rule can thus be implemented numerically as

Find  $\alpha\beta$  such that

$$\lambda_0^{(\alpha\beta)} + \sum_{i=1}^t \lambda_i^{(\alpha\beta)} m_i \leq \lambda_0^{(rs)} + \sum_{i=1}^t \lambda_i^{(rs)} m_i$$

as  $r$  varies from 0 to  $N$  and  $s$  varies from 1 to  $Q$ . Defining  $(N + 1)Q$  Lagrange multiplier vectors  $[\Lambda_{rs}]$  by

$$[\Lambda_{rs}] = \begin{bmatrix} \lambda_0^{(rs)} \\ \lambda_1^{(rs)} \\ \vdots \\ \lambda_t^{(rs)} \end{bmatrix}$$

and an augmented measurement vector as

$$[M] = \begin{bmatrix} 1 \\ m_1 \\ m_2 \\ \vdots \\ m_t \end{bmatrix}$$

allows the detection algorithm to be compactly expressed as a dot product operation.

Find  $\alpha\beta$  such that

$$\left[ \Lambda_{\alpha\beta} \right]^T [M] = \left[ \Lambda_{rs} \right]^T [M]$$

when compared to all  $(N + 1)Q$  lambda vectors.

The detection algorithm presented here is numerically attractive since all  $(N + 1)Q$  lambda vectors for the template densities can be precomputed. The only on-line computations required then are the information function expected value measurements and  $(N + 1)Q$  vector multiplications or dot products.

## Chapter IV. Information Functions

### Introduction

The goal of the minimum cross-entropy detection algorithm is the identification of objects contained in information cells independent of their position and orientation within the cell. To meet this goal and complete the detection rule definition, a set of orthonormal image moments will be developed and referenced to a standard coordinate system to completely define the information functions  $f_j(x,y)$ . The number and form of the information functions will then partly determine the accuracy and resulting cross-entropy  $H(e, t^{(k)})$  distances between the approximate scene density  $e(x,y)$  and the  $(N + 1)Q$  template densities  $t^{(k)}(x,y)$ .

### Image Moments

The concept of moments is used extensively in classical mechanics and statistics. In this dissertation, the two-dimensional  $(r + s)$ th order raw moments of the normalized information cell irradiance distribution  $q(x,y)$  are defined in terms of Riemann integrals

$$u_{rs} = \iint_{\text{cell}} x^r y^s q(x,y) dx dy$$

The irradiance distribution is a bounded function that can

have nonzero values only in a finite part of the  $xy$  plane. Because of these irradiance distribution function characteristics, moments of all orders exist and the double moment sequence  $\{u_{rs}\}$  is uniquely determined by the density  $q(x,y)$  and conversely  $q(x,y)$  is uniquely determined by  $\{u_{rs}\}$  (Hu, 1962).

The low-order moments can be used to define a standard coordinate system about which the moment sequence will be invariant. With the "target" as the predominant feature in the information cell where the term "target" also models a clutter configuration with the target of interest, this standard coordinate system will be invariant to changes in intensity, orientation and location of the "target." The zero-order moment is given by

$$u_{00} = \iint_{\text{cell}} q(x,y) dx dy$$

and represents the total image power. The image power is normalized to one as required of a probability density and this also provides a standard density that is invariant to uniform intensity variations. The first-order moments

$$u_{10} = \iint_{\text{cell}} xq(x,y) dx dy$$

and

$$u_{01} = \iint_{\text{cell}} yq(x,y) dx dy$$

can be used to define central moments  $\{v_{rs}\}$  which are invariant to translation of the "target" within the information cell. These first-order moments locate the centroid of the image irradiance distribution, i.e.  $\bar{x} = u_{10}/u_{00}$ ,  $\bar{y} = u_{01}/u_{00}$  and the central moments are then defined about the centroid as

$$v_{rs} = \iint_{\text{cell}} (x - \bar{x})^r (y - \bar{y})^s q(x,y) dx dy$$

From the definition of central moments it is easy to express the central moments in terms of the raw moments. For example, the first three moment orders are related by

$$\text{Zero Order:} \quad v_{00} = u_{00}$$

$$\text{First Order:} \quad v_{10} = u_{10} - \bar{x}u_{00} = 0$$

$$v_{01} = u_{01} - \bar{y}u_{00} = 0$$

$$\text{Second Order:} \quad v_{20} = u_{20} - \bar{x}^2 u_{00}$$

$$v_{11} = u_{11} - \bar{x}\bar{y}u_{00}$$

$$v_{02} = u_{02} - \bar{y}^2 u_{00}$$



A general formula for calculating the central moments in terms of the raw moments can be found using the binomial expansion  $(a + b)^n = \sum_{j=0}^n \binom{n}{j} a^{n-j} b^j$ . The resulting general expression for converting raw moments into central moments is

$$v_{rs} = \sum_{j=0}^r \sum_{k=0}^s \binom{r}{j} \binom{s}{k} (-x)^{r-j} (-y)^{s-k} u_{jk}$$

where the notation  $\binom{a}{b}$  denotes the usual binomial coefficient and equals  $a!/b!(a-b)!$ . Kanazawa also provides a FORTRAN program (Center) to calculate two-dimensional central moments from a set of two-dimensional raw moments using an alternate iterative relationship (Kanazawa, 1980:106).

Using second order moments, a second image invariant, the angle of minimum moment of inertia  $\theta$  can be used in addition to the center of mass given by the centroids. The quantities  $x, y$  and  $\theta$  together define an invariant reference frame for any information cell. In terms of raw moments the angle of minimum moment of inertia is defined by

$$\theta = \frac{1}{2} \tan^{-1} \frac{2(u_{00}u_{11} - u_{10}u_{01})}{(u_{00}u_{20} - u_{10}^2) - (u_{00}u_{02} - u_{01}^2)}$$

and defines a region's orientation within a two-fold degeneracy. The use of central moments converts the image invariants into the more intuitive concept of an invariant image ellipse. The second-order central moments

$$V_{20} = \iint_{\text{cell}} (x - \bar{x})^2 q(x,y) dx dy$$

$$V_{11} = \iint_{\text{cell}} (x - \bar{x})(y - \bar{y}) q(x,y) dx dy$$

$$V_{02} = \iint_{\text{cell}} (y - \bar{y})^2 q(x,y) dx dy$$

characterize the size and orientation of the image. Using only central moments up through second order, the original image can not be discriminated from a constant irradiance ellipse having definite size, orientation and eccentricity while centered at the image centroid (Teague, 1980). The semi-major axis  $x'$  and the semi-minor axis  $y'$  of the ellipse are shown in Figure 4.1 and define the principal axes of the pattern. Moments defined using the principal axes of the pattern are invariant to rotation and translation. Using central moments, the angle of minimum moment of inertia  $\theta$  reduces to the angle  $\phi$  that defines a rotation from the original  $x$  axis to the semi-major axis  $x'$  of the image ellipse. The tilt angle  $\phi$  is defined by

$$\phi = \frac{1}{2} \tan^{-1} \left[ \frac{2V_{11}}{V_{20} - V_{02}} \right]$$

where  $-\pi/2 \leq \tan^{-1}(x) \leq \pi/2$ . There is an ambiguity in the tilt angle  $\phi$  which can be resolved by selecting  $\phi$  as the angle between the  $x$  axis and the semi-major axis of the ellipse or as defined in Figure 4.1, having the image para-

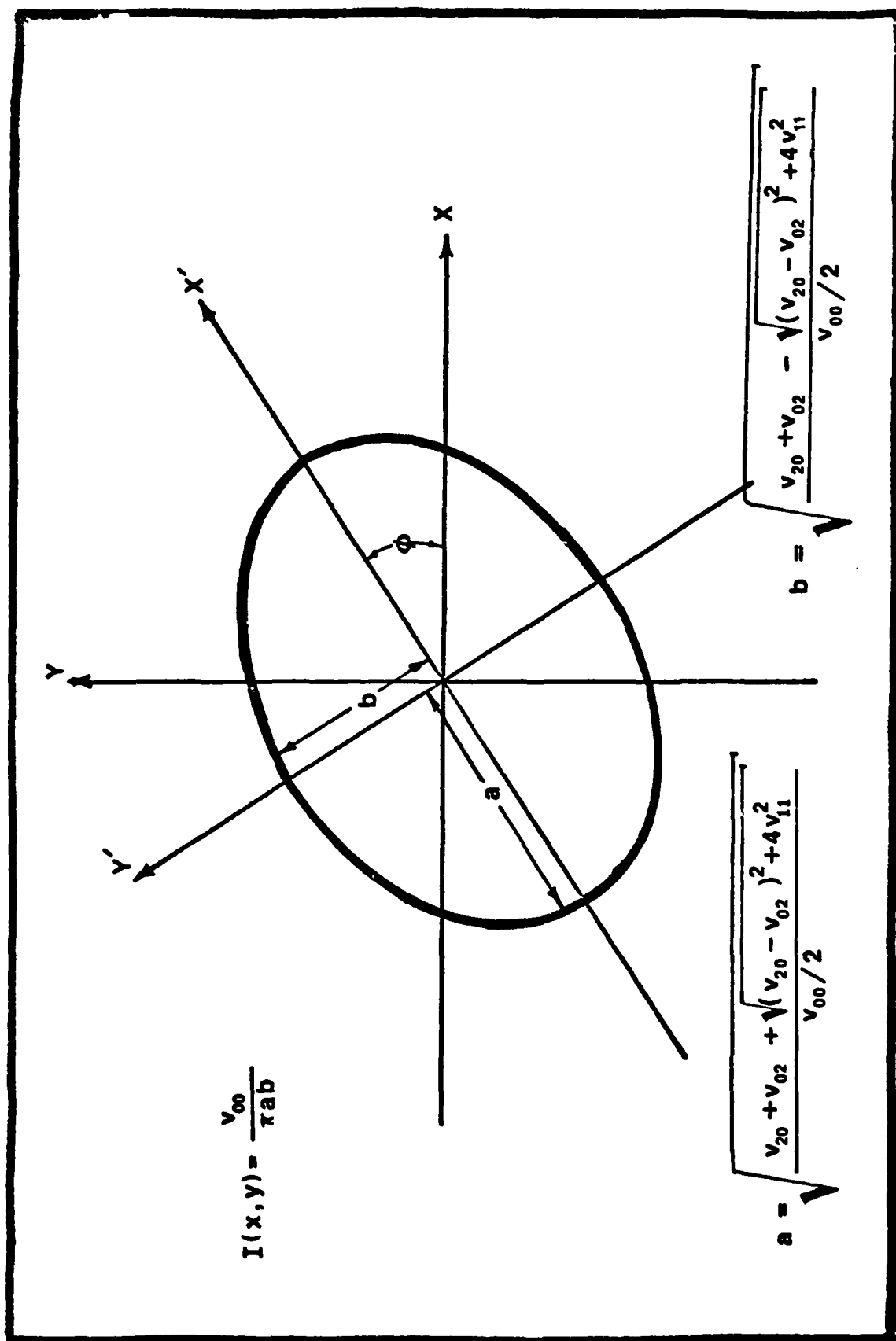


Fig. 4.1. Image Ellipse

meter a always greater than or equal to the parameter b.

The rotation to the invariant principal axes of the pattern corresponds to the orthogonal transformation

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

Using this orthogonal transformation a general expression for invariant moments can be defined in terms of a set of central moments  $\{V_{rs}\}$  as

$$w_{rs} = \iint_{\text{cell}} (x \cos \theta + y \sin \theta)^r (-x \sin \theta + y \cos \theta)^s q(x, y) dx dy$$

Again using the binomial expansion results in the expression

$$w_{rs} = \iint_{\text{cell}} \sum_{j=0}^r \binom{r}{j} (x \cos \theta)^{r-j} (y \sin \theta)^j \sum_{k=0}^s \binom{s}{k} (-x \sin \theta)^{s-k} (y \cos \theta)^k q(x, y) dx dy$$

which is equivalent to

$$w_{rs} = \sum_{j=0}^r \sum_{k=0}^s (-1)^{s-k} \binom{r}{j} \binom{s}{k} (\cos \theta)^{r-j+k} (\sin \theta)^{j+s-k} V_{r-j+s-k, j+k}$$

The general transformation expression shows that the set of central moments  $\{V_{rs}\}$  of order  $N = r + s$  transform into the

set of invariant moments  $\{w_{rs}\}$  of the same order  $N = r + s$ .

In summary then the raw moments  $u_{rs}$  of order two and below have been used to construct an invariant reference frame called the principal axes of the pattern which is invariant to uniform intensity variations, translation and rotation of the "target" within the information cell. General expressions were obtained for computing the invariant moments of any order from a set of raw moments of the same order. The raw image moments are first converted to central moments and the central moments are then mapped into invariant moments referenced to the principal axes of the pattern. The translation and rotation of the set of raw moments will be much faster numerically than translating and rotating the complete image before computing the set of invariant moments. Figure 4.2 illustrates the two equivalent methods of obtaining the desired set of invariant moments.

#### Orthonormal Moments

From functional analysis it is well-known that the general definition of the moment operator

$$w_{jk} = \iint_{\text{cell}} q(x,y) x^j y^k dx dy$$

has the form of a projection of the normalized irradiance function  $q(x,y)$  onto the subspace of monomials  $\{x^j y^k\}$ . The Weierstrass approximation theorem shows that the monomials form a complete basis set for a series expansion of  $q(x,y)$

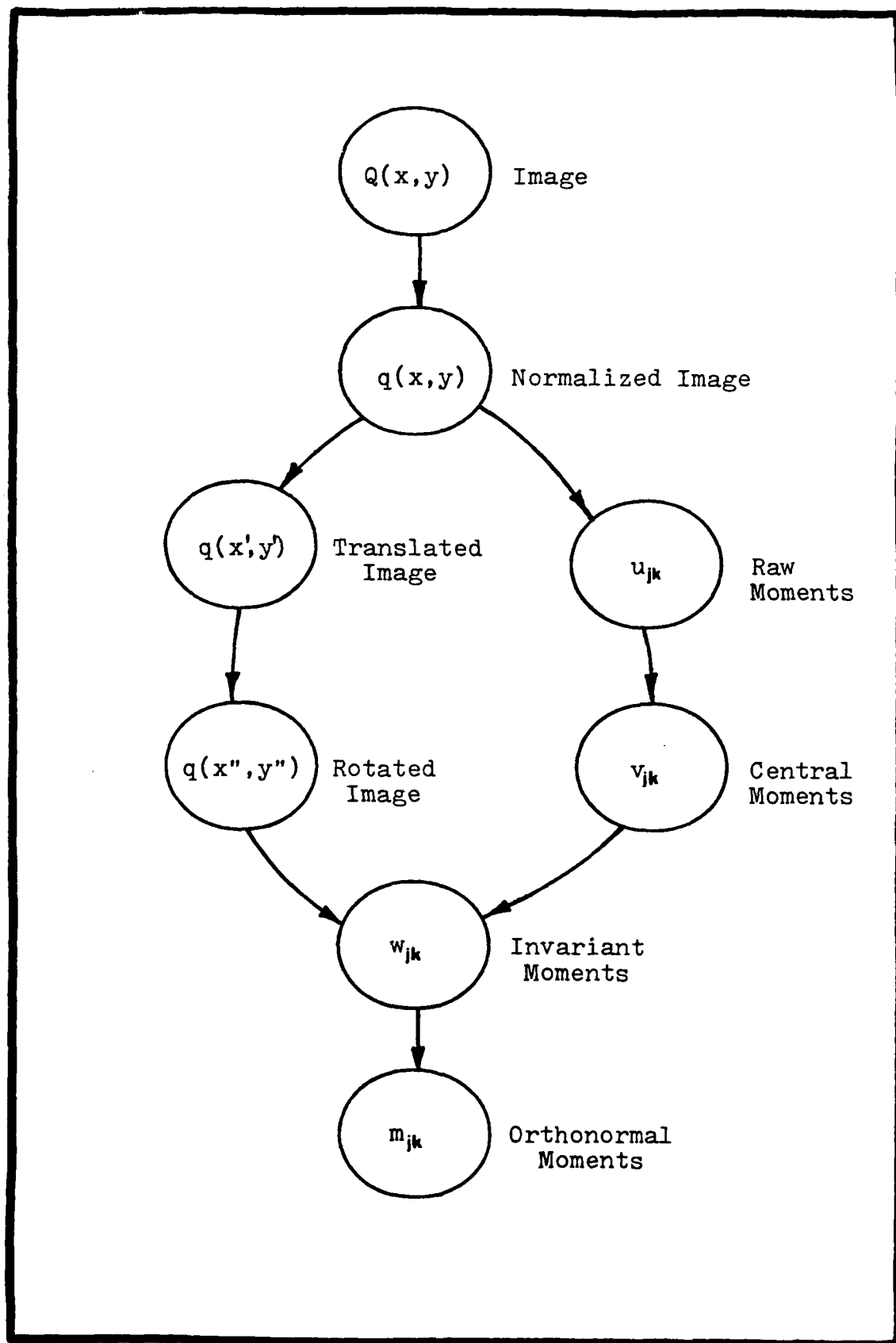


Fig. 4.2. Alternate Methods to Compute Invariant Moments

but the monomials do not form an orthogonal basis set. Using the Gram-Schmidt process on the linearly independent set  $\{1, x, x^2, \dots\}$  produces a well-known orthonormal set  $\{O_n(x)\}$  over the interval  $-1 \leq x \leq 1$  that is useful in constructing a two-dimensional orthonormal set based on the monomials (Kreyszig, 1978:176). The orthonormal elements have the general form

$$O_n(x) = \sqrt{\frac{2n+1}{2}} P_n(x)$$

where  $P_n(x)$  is the Legendre polynomial of order  $n$ . Using the orthonormal elements  $O_n(x)$  to produce information functions as linear combinations of the monomials allows  $q(x, y)$  to simultaneously have both of the following orthogonal series expansions:

$$q(x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \beta_{mn} O_m(x) O_n(y)$$

and

$$q(x, y) = \exp \left[ \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \lambda_{mn} O_m(x) O_n(y) \right]$$

This model of the image density function is an extension of the model developed by Neyman (Neyman, 1937) and used in several articles by Crain (Crain, 1977, 1974, 1976) dealing with approximating univariate probability densities. The

Neyman model for the unknown information cell density function  $q(x,y)$  corresponds to the minimum cross-entropy (maximum entropy) density

$$q(x,y) = \exp \left[ - \sum_{i=0}^{\infty} \lambda_i f_i(x,y) \right]$$

with an infinite number of information functions  $f_i(x,y)$  formed as a product of normalized Legendre polynomials. The Neyman infinite series expansion expressed as a product of normalized Legendre polynomials is basically a summation by infinite rows of a matrix of series terms and must be approximated to be of any practical value. The approximation is obtained by summing along finite diagonals and truncating at a finite order  $N_{\max}$  to obtain the expressions:

$$q(x,y) \cong \sum_{i=0}^{N_{\max}} \sum_{n=0}^i \beta_{i-n,n} O_{i-n}(x) O_n(y)$$

and

$$q(x,y) \cong \exp \left[ \sum_{i=0}^{N_{\max}} \sum_{n=0}^i \lambda_{i-n,n} O_{i-n}(x) O_n(y) \right]$$

This result is used by Teague (Teague, 1980) and is the basic equation required to approximate the unknown information cell density  $q(x,y)$ . The approximation for the information cell density also corresponds to a minimum cross-entropy (maximum entropy) density with  $(N_{\max} + 1)(N_{\max} + 2)/2$  information



functions  $f_j(x,y)$  formed as products of the normalized Legendre polynomials. Due to the rotational properties of moments, all moments of a given order must be included in the series expansion and treated as having equal importance when constructing an approximate density function. The number of information functions required in an approximate density function of order  $N_{\max}$  is shown in Figure 4.3 with each diagonal line in the figure corresponding to a different moment order beginning with one function and zero order. The accuracy of the truncated approximation for  $q(x,y)$  improves as  $N_{\max}$  is increased, however, the numerical difficulties of solving a large system of nonlinear equations for  $[\Lambda]$  also increases as the size of the corresponding lambda vector grows with  $N_{\max}$ . The selection of the moment order  $N_{\max}$  is thus a compromise between accuracy as measured by cross-entropy  $H(q, t^{(rs)})$  and the numerical processing time required to compute the set of lambda vectors  $[\Lambda_{\alpha\beta}]$  and the augmented measurement vectors  $[M]$ .

The problem is then to select  $N_{\max}$  for the set of template densities  $t^{(rs)}(x,y)$ . The template density representation must be accurate enough to provide a small probability of error for the detection algorithm and also not require an excessive amount of processing time. Cross-entropy  $H(q, t^{(rs)})$  serves as an information theoretic distance between the true density  $q(x,y)$  and the model density  $t^{(rs)}(x,y)$ . Because  $H(q, t^{(rs)})$  is nonnegative and  $H(q, t^{(rs)}) = 0$  if and only if

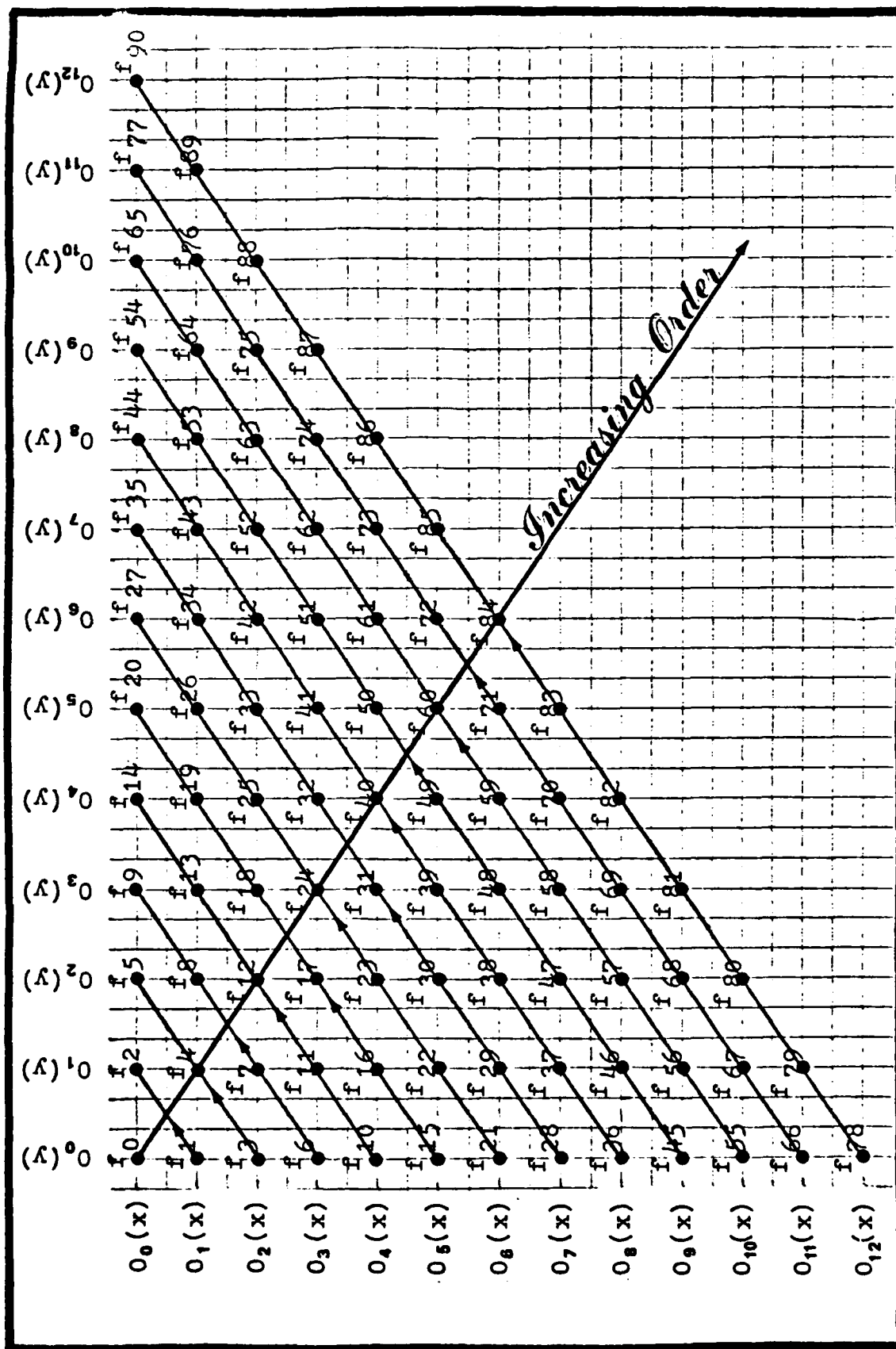


Fig. 4.3. Information Function Construction

$q(x,y) = t^{(rs)}(x,y)$  almost everywhere, cross-entropy between the model and the true density can be used to select  $N_{\max}$  the required model order. The cross-entropy distance is given by

$$H(q, t^{(rs)}) = \iint_{\text{cell}} q(x,y) \ln \left[ \frac{q(x,y)}{t^{(rs)}(x,y)} \right] dx dy$$

Substituting the Neyman and minimum cross-entropy density forms gives the expression

$$H(q, t^{(rs)}) = \iint_{\text{cell}} q(x,y) \ln \left[ \frac{\exp - \sum_{i=0}^{\infty} \beta_i f_i(x,y)}{\exp - \sum_{n=0}^{\infty} \alpha_n f_n(x,y)} \right] dx dy$$

Rearranging and taking expected values results in

$$H(q, t^{(rs)}) = \sum_{i=0}^t (\alpha_i - \beta_i) m_i - \iint_{\text{cell}} q(x,y) \left[ \sum_{n=t+1}^{\infty} \beta_n f_n(x,y) \right] dx dy$$

which is known to approach zero as  $t$  approaches infinity. With the infinite sum in the last expression converging to some function  $Z(x,y)$  the cross-entropy distance measure will take the form

$$H(q, t^{(rs)}) = \sum_{i=0}^t (\alpha_i - \beta_i) m_i + C(t)$$

where  $C(t)$  is a constant for each value of  $t$  in the template density expression. Since this expression for cross-entropy can not be evaluated analytically, in Figure 6.1

$H(q, t^{(rs)})$  is numerically evaluated for a block tank template and plotted as a function of  $t$ , the number of information functions used in the template. In general, for the minimum cross-entropy detection algorithm to be effective the sum of the expected values of the high order terms or the constant  $C(t)$  must convey a small amount of information. The information content of the high order terms in turn depends on the number and magnitude of the variations in the density that is being approximated by the template. Because of this target dependence,  $H(q, t^{(rs)})$  is evaluated numerically in Chapter VI using a template with many abrupt and relatively large changes in the density function to provide an approximate worst case relationship between the number of information functions and the resulting cross-entropy  $H(q, t^{(rs)})$ .

#### Legendre Polynomials

To complete the information function definition, we need explicit expressions for the Legendre polynomials that are used to form  $f_n(x, y)$ . These orthogonal polynomials are defined over the interval  $[-1, 1]$  and have the general explicit expression

$$P_n(x) = \frac{1}{2^n} \sum_{m=0}^{\lfloor n/2 \rfloor} (-1)^m \binom{n}{m} \binom{2n-2m}{n} x^{(n-2m)}$$

where  $\lfloor n/2 \rfloor$  denotes the greatest integer not exceeding  $n/2$ . Legendre polynomials also satisfy the following recurrence relation (Courant, 1953:86):

$$P_n(x) = \left(\frac{2n-1}{n}\right)xP_{n-1}(x) - \left(\frac{n-1}{n}\right)P_{n-2}(x)$$

Table I explicitly defines the first thirteen Legendre polynomials and the required normalization factor  $\sqrt{(2n+1)/2}$ . The Legendre polynomials shown can be used to define the 91 information functions required by Figure 4.3 for a twelfth order approximation of the true density function. Figure 4.3 also shows how all 91 information functions are constructed from the normalized Legendre polynomials. For example, the 44th information function shown as number 43 on the figure is given by

$$f_{43}(x,y) = O_1(x) \cdot O_7(y)$$

and using the expressions given in Table I becomes

$$f_{43}(x,y) = (1.22x)(2.74)(26.81y^7 - 43.31x^5 + 19.69y^3 - 2.19y)$$

or

$$f_{43}(x,y) = 89.31xy^7 - 144.66xy^5 + 65.76xy^3 - 7.31xy$$

The expected value of this information function can be written in terms of invariant moments as

$$m_{43} = 89.31w_{17} - 144.66w_{15} + 65.76w_{13} - 7.31w_{11}$$

Table I. Legendre Polynomials

Symbol	$\sqrt{\frac{2n+1}{2}}$	Explicit Expression
$P_0(x)$	0.707	1
$P_1(x)$	1.22	x
$P_2(x)$	1.58	$1.5x^2 - 0.5$
$P_3(x)$	1.87	$2.5x^3 - 1.5x$
$P_4(x)$	2.12	$4.38x^4 - 3.75x^2 + 0.38$
$P_5(x)$	2.35	$7.88x^5 - 8.75x^3 + 1.88x$
$P_6(x)$	2.55	$14.44x^6 - 19.69x^4 + 6.56x^2 - 0.31$
$P_7(x)$	2.74	$26.81x^7 - 43.31x^5 + 19.69x^3 - 2.19x$
$P_8(x)$	2.92	$50.27x^8 - 93.85x^6 + 54.15x^4 - 9.85x^2 + 0.27$
$P_9(x)$	3.08	$94.96x^9 - 201.09x^7 + 140.75x^5 - 36.09x^3 + 2.45x$
$P_{10}(x)$	3.24	$180.42x^{10} - 427.31x^8 + 351.90x^6 - 117.31x^4 + 13.53x^2 - 0.24$
$P_{11}(x)$	3.39	$344.42x^{11} - 902.05x^9 + 854.57x^7 - 351.88x^5 + 58.64x^3 - 2.69x$
$P_{12}(x)$	3.54	$660.25x^{12} - 1894.68x^{10} + 2030.05x^8 - 997.24x^6 + 219.98x^4 - 17.57x^2 + 0.22$

The measurement vector  $[M]$  required in the detection algorithm is then completely defined by the set of invariant moments  $\{w_{mn}\}$  .

In summary, a set of raw moments  $\{u_{mn}\}$  is produced and converted into central moments  $\{v_{mn}\}$  about the pattern centroid. The central moments are then rotated and become the set of invariant moments  $\{w_{mn}\}$  about the principal axes of the pattern. The set of orthonormal moments  $\{m_r\}$  that form the measurement vector  $[M]$  are then computed as linear combinations of these invariant moments. Given these definitions, the detection algorithm is ready to process information cells.

## Chapter V. Numerical Techniques and Performance Analysis

The key to implementing the minimum cross-entropy detection algorithm is the ability to find the correct Lagrange multiplier vector for a given set of measurements. The problem mathematically is to find  $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_t)^T$  such that

$$[F(\Lambda)] = \begin{bmatrix} F_0(\Lambda) \\ F_1(\Lambda) \\ \vdots \\ F_t(\Lambda) \end{bmatrix} = \begin{bmatrix} \iint e(x,y) dx dy - 1 \\ \iint f_1(x,y) e(x,y) dx dy - m_1 \\ \vdots \\ \iint f_t(x,y) e(x,y) dx dy - m_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where  $e(x,y) = p(x,y) \exp\{-\lambda_0 - \lambda_1 f_1(x,y) - \dots - \lambda_t f_t(x,y)\}$  is the minimum cross-entropy density with a uniform prior density. The  $(t + 1)$  constraints are nonlinear and except for a few restricted cases cannot be solved directly for the lambda vector. Several authors discuss iterative numerical schemes for simultaneous solution of a system of nonlinear equations. Johnson (Johnson, 1979b:24) provides a computer program written in APL for solving discrete cross-entropy minimization problems with arbitrary positive priors that is based on the Newton-Raphson method. Gokhale and Kullback



(Gokhale, 1978) describe a somewhat different algorithm also based on the Newton-Raphson method that has been implemented in PL/1. Agmon, Alhassid and Levine (Agmon, 1979:250) describe yet another discrete cross-entropy minimization algorithm using a uniform prior and a FORTRAN implementation. Miller also provides an alternate FORTRAN implementation in one-dimension that is based on the Newton-Raphson method (Miller, 1980:45).

The Newton method is an iterative scheme based on the relationship:

$$[\Delta\Lambda] = [\Lambda^{(n)}] - [\Lambda^{(n+1)}] = [J]^{-1} [F(\Lambda^{(n)})]$$

where  $[\Lambda^{(n)}]$  is the Lagrange multiplier vector lambda for the nth iteration and  $[J]$  is the Jacobian matrix for  $[F(\Lambda^{(n)})]$ . The initial estimate  $[\Lambda^{(0)}]$  is selected and the equation solved for  $[\Lambda^{(1)}]$ . The procedure repeats for  $[\Lambda^{(2)}]$ ,  $[\Lambda^{(3)}]$ , ...,  $[\Lambda^{(n)}]$ ,  $[\Lambda^{(n+1)}]$  until the difference  $[\Delta\Lambda]$  is less than a predefined value which insures convergence has occurred. The equation that must be solved numerically to implement the Newton method requires an evaluation of the Jacobian matrix  $[J]$  during every iteration for a new lambda vector. The Jacobian matrix has terms of the form  $\partial F_i(\Lambda)/\partial \lambda_j$  and  $[J]$  is then a  $(t + 1) \times (t + 1)$  symmetric matrix. Convergence and rate of convergence of the Newton algorithm are dependent on the initial estimate  $[\Lambda^{(0)}]$ . Many authors address the

theoretical convergence criteria of the Newton method (see (Ortega, 1970) and (Saaty, 1964)) which in general define a neighborhood about the solution vector where convergence is assured if the initial estimate falls in this neighborhood.

The Newton method has thus been used almost exclusively in solution schemes appearing in the technical literature. Past applications, however, worked with a small number of constraints and have encountered problems with ill-conditioning in the computer generated Jacobian matrix and selection of an appropriate initial estimate of the lambda vector (Miller, 1980:50). Using 91 constraints accentuates these numerical problems to the point that the second-order Newton method (Dodes, 1978) must be abandoned for lower order methods that do not use derivative information.

#### Lambda Vector Solution

The processing required to find the lambda vector is the main burden of the minimum cross-entropy detection algorithm. The lambda vector for each template density is, however, pre-computed and stored for use in the detection algorithm. A zero order method was selected to solve for the lambda vector since the procedure must only be accomplished once for each template density and most numerical problems are avoided. The Cyclic Coordinate Method (Bazaraa, 1979:271) is a multi-dimensional search procedure that does not use derivatives. The only required information is that  $\Lambda \in L$  where  $L$  has the form  $L = \{\Lambda : a_i \leq \lambda_i \leq b_i\}$ . The search procedure requires

only a defined search interval and has been implemented on the Data General Eclipse S/250 Integral Array Processor. When performed on the array processor, the cyclic coordinate search method produces results faster than a comparably dimensioned Newton algorithm written in FORTRAN. The search problem is then given the vector of constraint relationships

$$F_i(\lambda) = \iint_{\text{cell}} f_i(x, y) p(x, y) \exp \left[ - \sum_{u=0}^t \lambda_u f_u(x, y) \right] dx dy - m_i = 0$$

$$(i = 1 \dots t)$$

find the lambda vector required to define  $e(x, y)$ . The discrete approximation of this equation can then be written in the form

$$\sum_{i=1}^n \sum_{m=1}^n \left\{ f_i(x_i, y_m) - m_i \right\} p(x_i, y_m) \exp \left[ - \sum_{u=0}^t \lambda_u f_u(x_i, y_m) \right] = 0$$

$$(i = 1 \dots t)$$

Using a uniform prior density and then canceling terms results in the equivalent expression:

$$\sum_{i=1}^n \sum_{m=1}^n \left\{ f_i(x_i, y_m) - m_i \right\} \exp \left[ - \sum_{u=1}^t \lambda_u \left[ f_u(x_i, y_m) - m_u \right] \right] = 0$$

$$(i = 1 \dots t)$$

The  $t = 91$  equations defining the Lagrange parameter vector  $\Lambda$  are implicit and nonlinear. The direct numerical solution of this vector constraint equation would be computationally cumbersome. However, using a result derived by Agmon et al (Agmon, 1979), this problem can be recast as a simpler variational problem. The technique requires that a "potential" function which is concave for any trial set of Lagrange parameters be defined. The values of the Lagrange multiplier parameters can then be determined as the set which minimizes the potential. Agmon et al provide the following lemma which has direct application to the nonlinear constraint equation given above:

Let  $\Omega \subset \mathbb{R}^l$  be a simply connected domain. Let  $F: \Omega \rightarrow \mathbb{R}^l$  be a continuously differentiable vector function. Denote its Jacobian by  $J$ , that is  $J_{ij} = \partial F_i / \partial \lambda_j$  and suppose it is a symmetric positive definite matrix. The problem of solving the set of nonlinear equations  $F(\Lambda) = 0$  is equivalent to finding a minimum of a concave scalar potential function  $\mathcal{F}$ .

The solution of the system of nonlinear equations  $F(\Lambda) = 0$  is then found to be equivalent to minimizing the following scalar potential function

$$\mathcal{F}(\Lambda) = \ln \left\{ \sum_{i=1}^n \sum_{m=1}^n \exp \left[ - \sum_{k=1}^l \lambda_k (f_k(x_i, y_m) - m_k) \right] \right\}$$

The solution program Lambda assumes  $L = \{\lambda: -25 \leq \lambda_i \leq 25\}$  and has always converged to a minimum for  $\mathcal{F}(\lambda)$ . Since the potential function is concave, should a component of the solution vector lie outside this interval the algorithm will select a value of  $\pm 25$  and signal that the search interval should be expanded. The functional minimization program uses a  $32 \times 32$  grid to compute  $\mathcal{F}(\lambda)$  and a sequence of decreasing search intervals to reach the solution vector.

The final component of the solution vector  $\lambda_0$  is found from the requirement that the resulting minimum cross-entropy density  $e(x,y)$  integrate to one. The result then

$$\lambda_0 = \ln \left[ \iint_{\text{cell}} \exp \left( - \sum_{k=1}^i \lambda_k f_k(x,y) \right) dx dy \right]$$

To implement the functional minimization scheme and produce image moments requires an effective quadrature scheme. The quadrature algorithm used in all programs will be developed in the next section.

### Numerical Quadrature

The first Newton-Cotes formula known as the Trapezoidal Rule gives a relationship which forms the basis of an effective quadrature scheme. In one dimension with the interval of integration divided in  $n$  parts the Trapezoidal Rule states that (Young, 1972:371):

$$\int_a^b f(x) dx = h \left[ \frac{1}{2}(f_0 + f_n) + \sum_{k=1}^{n-1} f_k \right] - \frac{nh^3}{12} f''(\xi), \xi \in (a,b)$$

where  $f_k = f(a + kh)$  and  $h = (b - a)/n$ . The first processing requirement for this quadrature scheme is the production of a set of raw moments  $\{u_{pq}\}$  from the unknown image density function  $q(x,y)$  or template density  $t(x,y)$ . Raw moments are then transformed into orthonormal Legendre moments  $[M]$  for use in the detection algorithm or the iterative cyclic coordinate method.

The two-dimensional raw moment  $u_{pq}$  is defined in Cartesian coordinates as

$$u_{pq} = \int_{-c}^c x^p \left[ \int_{-c}^c y^q q(x,y) dy \right] dx$$

which can be rewritten as

$$u_{pq} = \int_{-c}^c x^p d_q(x) dx$$

where

$$d_q(x) = \int_{-c}^c y^q q(x,y) dy$$

The one dimensional Newton-Cotes formula will be used to approximate  $d_q(x)$ . Using only the first term of the Trapezoidal Rule gives

$$d_q(x_i) \approx h \left[ \frac{1}{2} \left( y_0^q q(x_i, y_0) + y_n^q q(x_i, y_n) \right) + \sum_{j=1}^{n-1} y_j^q q(x_i, y_j) \right]$$

The approximation can now be applied again to produce the raw moments  $u_{pq}$ . The raw moments are then:

$$u_{pq} \approx h \left[ \frac{1}{2} (x_0^p d_q(x_0) + x_n^p d_q(x_n)) + \sum_{k=1}^{n-1} x_k^p d_q(x_k) \right]$$

Completely expanding this approximation gives a clearer picture of the factors involved in  $u_{pq}$ . The expanded approximation for the raw moments  $u_{pq}$  is:

$$\begin{aligned} u_{pq} \approx & \frac{h^2}{4} \left[ x_0^p y_0^q q(x_0, y_0) + x_0^p y_n^q q(x_0, y_n) + x_n^p y_0^q q(x_n, y_0) + x_n^p y_n^q q(x_n, y_n) \right] \\ & + \frac{h^2}{2} \left[ x_0^p \sum_{j=1}^{n-1} y_j^q q(x_0, y_j) + x_n^p \sum_{k=1}^{n-1} y_k^q q(x_n, y_k) + y_0^q \sum_{i=1}^{n-1} x_i^p q(x_i, y_0) \right. \\ & \left. + y_n^q \sum_{m=1}^{n-1} x_m^p q(x_m, y_n) \right] + h^2 \sum_{k=1}^{n-1} \sum_{j=1}^{n-1} x_k^p y_j^q q(x_k, y_j) \end{aligned}$$

The first grouping of terms in the expression represents the contribution from the four corner points of the sampled unknown (or template) density array  $q(x_i, y_i)$  (or  $t(x_i, y_i)$ ). The second grouping of terms represents the contribution from the remaining "edge" sample points. The last term then represents all the interior sample points and when using a 256 x 256 sampled density array represents 98.44% of the possible contribution to the moments. When a white border is used with the density matrix the "edge" terms will make no contribution at all to the moment approximation and can be ignored. With these insights the raw moment approximation

used in this dissertation will be

$$u_{pq} \approx h^2 \sum_{k=0}^n \sum_{j=0}^n x_k^p y_j^q q(x_k, y_j)$$

For computer computation using this approximation all moments through order  $L$  can be computed with two matrix multiplications. The matrix equation for the raw moment matrix is:

$$[u_{pq}] = \begin{bmatrix} u_{00} & u_{01} & u_{02} & \cdots & u_{0L} \\ u_{10} & u_{11} & u_{12} & \cdots & u_{1L} \\ u_{20} & u_{21} & u_{22} & \cdots & u_{2L} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ u_{L0} & u_{L1} & u_{L2} & & u_{LL} \end{bmatrix} = \begin{matrix} \\ \\ \\ \\ \\ (L+1) \times (L+1) \end{matrix}$$

$$h^2 \begin{bmatrix} x_0^0 & x_1^0 & x_2^0 & \cdots & x_n^0 \\ x_0^1 & x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ x_0^L & x_1^L & x_2^L & & x_n^L \end{bmatrix} \times \begin{bmatrix} q(x_0, y_0) & q(x_0, y_1) & \cdots & q(x_0, y_n) \\ q(x_1, y_0) & q(x_1, y_1) & \cdots & q(x_1, y_n) \\ q(x_2, y_0) & q(x_2, y_1) & \cdots & q(x_2, y_n) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ q(x_n, y_0) & q(x_n, y_1) & \cdots & q(x_n, y_n) \end{bmatrix}$$



$$x \begin{bmatrix} y_0^0 & y_0^1 & y_0^2 \dots y_0^L \\ y_1^0 & y_1^1 & y_1^2 \dots y_1^L \\ y_2^0 & y_2^1 & y_2^2 \dots y_2^L \\ \vdots & \vdots & \vdots & \vdots \\ y_n^0 & y_n^1 & y_n^2 & y_n^L \end{bmatrix}$$

With the Cartesian coordinate system origin centered in the bordered image density function and the image normalized to  $-1 \leq x, y \leq +1$  the subroutine moment in Figure 5.1 computes the normalized raw moment array.

#### Detection Algorithm Infrastructure

The basic software modules required in the minimum cross-entropy detection algorithm are shown in Figure 5.1. The diagram shows both on-line (solid line) and pre-processing (broken line) software modules with the interdependence between the two types of processing. When using the detection algorithm to process information cells, the  $(N + 1)Q$  lambda vectors  $(\lambda_0^{(k)}, \lambda_1^{(k)} \dots \lambda_i^{(k)})$  will be stored as constants for use in the detection program.

The pre-processing starts with a set of  $(N + 1)(Q)$  template densities used to "train" the detection algorithm. These template densities represent pure clutter and  $N$  targets of current interest all superimposed on the  $Q$  clutter backgrounds. The analog to digital program (A - D) produces

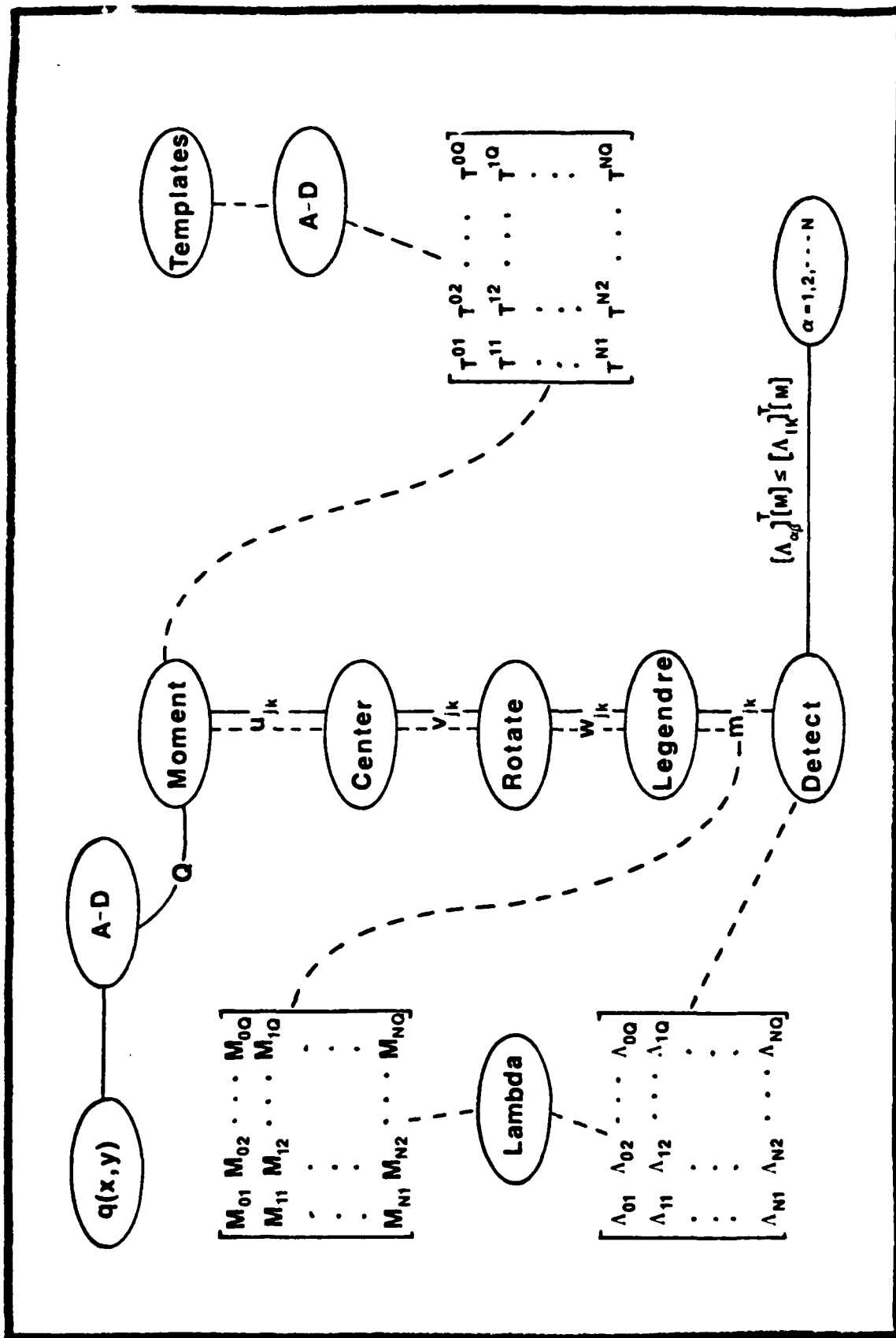


Fig. 5.1. Software Module Interdependence

$(N + 1)Q$  arrays of sample values each of dimension  $256 \times 256$ . The sample values within the arrays have also been quantized to 16 levels. Each array then contains 65,536 integer values in the range 0-15. The program moment produces a set of 91 raw moments  $\{u_{rs}\}$  for each template array. The programs center, rotate and Legendre map the moments for each template into Legendre moments about the principal axis. The final preprocessing program lambda iteratively solves for 91  $\lambda$ 's required to define the minimum cross-entropy density for each template and also used in the detection algorithm.

The on-line processing for each unknown information cell simply produces a set of 91 Legendre moments  $[M]$  using the programs outlined in the preprocessing section. With the Legendre moment vector and the  $(N + 1)Q$  lambda vectors the detection program produces  $(N + 1)Q$  dot products. The matrix of dot product values is searched for the smallest element and the row number of that element determines the classification decision for that information cell. The on-line processing is then repeated for each new information cell. The lambda values, however, remain fixed for all information cells presented to the detection algorithm and thus preprocessing is performed only once for a given set of template densities.

#### Performance Analysis

Given the theory and now the software for a minimum cross-entropy detection algorithm, an expected performance

analysis will complete the presentation of this detection rule. The next chapter will then explore actual performance on a set of test scenes and verify the impact of the salient factors influencing probability of error. The analysis in this section will attempt to use concepts from statistical communication theory to identify the major factors determining  $P(\epsilon)$  in the minimum cross-entropy algorithm. Statistical communication theory has as its goal the detection or estimation of signals in the presence of noise, but because of the difficulty of establishing useful statistical assumptions (Duda, 1973:324) it has found few applications in scene analysis.

With statistical communication theory techniques in mind, the information cell density can be simply modeled as the sum of two terms:

$$q(x,y) = s(x,y) + n(x,y)$$

The term  $s(x,y)$  represents the expected signal or the template density used to train the detection algorithm, i.e. the densities that correspond to the stored lambda vectors. The other component  $n(x,y)$  represents the clutter that was not expected nor modeled by the template densities and acts like a noise term to the detection rule. Looking at the binary decision case for simplicity, each of the two training densities has a set of moments and a corresponding precomputed

lambda vector. Then associated with the target template are  $[M_T]$  and  $[A_T]$  while the clutter template has associated vectors  $[M_C]$  and  $[A_C]$ .

Using the minimum cross-entropy decision rule will first require computing a set of moments over the information cell or

$$\iint_{\text{cell}} q(x,y) f_i(x,y) dx dy = m_i$$

$$(i = 0 \dots t)$$

Substituting in the additive noise model gives the expression

$$\iint_{\text{cell}} s(x,y) f_i(x,y) dx dy + \iint_{\text{cell}} n(x,y) f_i(x,y) dx dy = m_i$$

$$(i = 0 \dots t)$$

Written as a vector this expression becomes

$$[M_S] + [M_N] = [M]$$

The first moment vector  $[M_S]$  represents one of the two expected signals, i.e. target or clutter. The second term  $[M_N]$  is a noise perturbation vector caused by the unexpected clutter in the scene. The minimum cross-entropy detection algorithm templates are produced using a uniform prior den-

sity and thus have maximum entropy consistent with the moment constraints. Entropy has been related to scene structure (Watanabe, 1981) where structure refers to the confinement of scene energy to a small number of pixels and large transitions in pixel energy. The smaller the scene entropy the larger the scene structuredness. The minimum cross-entropy templates thus have as little structure as possible and still conform to the moment constraints. The normalized scene energy is smoothly spread over as many pixels as allowed by the moment constraints to produce the template densities. The minimum cross-entropy decision rule is thus inherently robust (Rey, 1978) to small moment perturbations which correspond to changes in the assumed underlying density. The changes in the underlying distribution have minimal impact on the detection algorithm since small perturbations are smoothed away in the process of constructing the maximum entropy templates.

The minimum cross-entropy decision rule then takes the perturbed information cell moment vector and performs a dot product operation with each of the stored template lambda vectors to produce

$$[\Lambda_c]^T[M] \underset{H_c}{\overset{H_t}{\gtrless}} [\Lambda_t]^T[M]$$

where  $H_t$  corresponds to selecting the target hypothesis and  $H_c$  corresponds to selecting the clutter hypothesis. Expanding the moment vector into its components gives

$$[\Lambda_C]^T[M_S] + [\Lambda_C]^T[M_N] \stackrel{H_T}{\geq} [\Lambda_T]^T[M_S] + [\Lambda_T]^T[M_N]$$

Given that the signal term corresponds to a target, the first term above will be a relatively large constant (K) while the third term will be a small constant (k). The results occur since both terms represent a cross-entropy that has been shown to be positive in all cases and very small for corresponding moment and lambda vectors. Therefore, the decision rule becomes

$$(K - k) \stackrel{H_T}{\geq} [\Lambda_T]^T[M_N] - [\Lambda_C]^T[M_N]$$

The terms on the right can be viewed as a particular realization of clutter from a large ensemble of possible clutter configurations and are each thus realizations of random variables. Then the decision rule can be written as

$$D = (K - k) \stackrel{H_T}{\geq} r_T - r_C = r$$

Since the random variable  $r$  is formed as the difference of similar random variables,  $r$  should have approximately zero mean and some variance  $\sigma^2$ . Type II errors are then made when  $r$  exceeds  $D$  and hypothesis  $H_C$  is declared to be true. Using Tchebycheff's Inequality gives an immediate probability of error expression for equally likely hypothesis as

$$P(\epsilon) = P(|r| > D) \leq \frac{\sigma^2}{D^2}$$

Because the random variables  $\underline{r}_T$  and  $\underline{r}_C$  are formed as a sum of approximately independent random variables as

$$r_T = \lambda_1^{(T)} m_1^{(N)} + \lambda_2^{(T)} m_2^{(N)} + \dots + \lambda_t^{(T)} m_t^{(N)}$$

and

$$r_C = \lambda_1^{(C)} m_1^{(N)} + \lambda_2^{(C)} m_2^{(N)} + \dots + \lambda_t^{(C)} m_t^{(N)}$$

the central-limit theorem (Papoulis, 1965:266) applies and  $\underline{r}$  will approach a Gaussian density as  $t$  becomes large. The probability of error can then be expressed as

$$P(\epsilon) = \frac{1}{2} - \text{erf}\left(\frac{D}{\sigma}\right)$$

where

$$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du$$

Now given the ratio  $D/\sigma$  a much better estimate of  $P(\epsilon)$  can be achieved with this approximation than the upper bound provided by Tchebycheff's Inequality. Figure 5.2 shows how  $P(\epsilon)$  varies with  $D/\sigma$ .

Both approaches to  $P(\epsilon)$  have shown a dependence on the ratio  $\sigma/D$  in predicting the minimum cross-entropy detection algorithm expected performance. The requirements for high



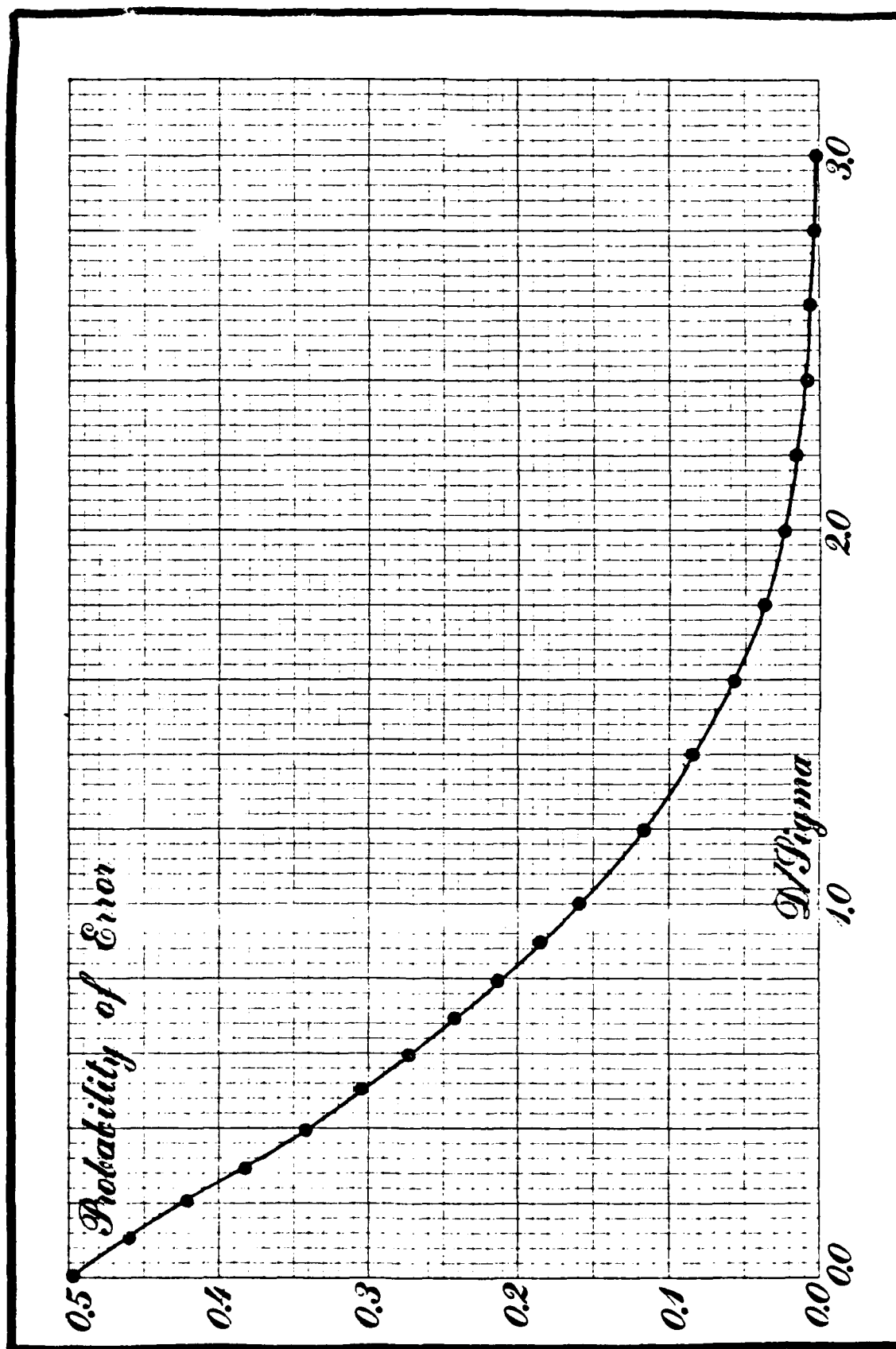


Fig. 5.2. Probability of Error versus Correlation Difference over Sigma

performance are to keep the clutter standard deviation sigma as small as possible and generate a large correlation difference D. The factors which combine to determine detection performance are now seen to be the number of moments used to characterize the target, the relative target to scene area and the amount of clutter modeled with the target in the template density. Increasing the number of information functions and thus the number of moments used to characterize the target will increase D since the k term will approach zero as more functions are used in the minimum cross-entropy template. The variance sigma squared can be decreased by reducing the scene/target ratio. The increased relative target size will then allow smaller variation in the clutter field since more of the scene will be represented by the s(x,y) term of the additive model. The variance can also be decreased by having more training densities with probable clutter configurations built into the target model. Again the signal term will account for more of the clutter and reduce the possible variance associated with the noise term n(x,y) in the additive model.

The clutter variance, however, can not be measured since the ensemble of clutter fields is not known and therefore, this is not a practical method of projecting the detection algorithm performance. Cross-entropy  $H(q,t)$  defined as

$$H(q,t) = \iint_{\text{cell}} q(x,y) \ln \left[ \frac{q(x,y)}{t(x,y)} \right] dx dy$$

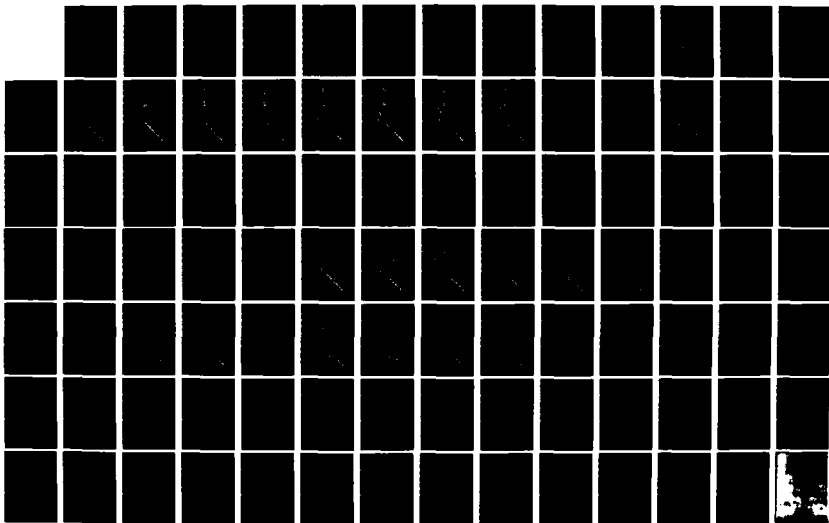
AD-A124 975

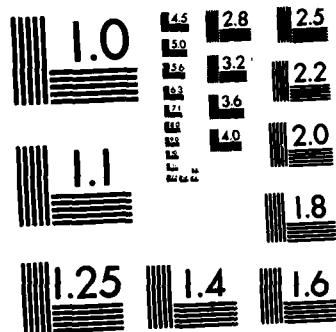
INFORMATION THEORETIC DETECTION OF OBJECTS EMBEDDED IN  
CLUTTERED AERIAL SCENES(U) AIR FORCE INST OF TECH  
WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI... F D TILLER  
JUL 82 AFIT/DS/EE/82-1 F/G 14/5

2/2

UNCLASSIFIED

NL





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

provides an alternate method of predicting the detection rule performance. The cross-entropy  $H(q,t) = H(q,p) - H(t,p)$  serves as a measure of the distance between the scene density  $q(x,y)$  and the template density  $t(x,y)$ . Minimizing  $H(q,t)$  has the effect of minimizing the number of pixel differences between the training template and the scene density and therefore limits the magnitude of the clutter variance sigma squared. Minimizing the template to scene cross-entropy  $H(q,t)$  is thus equivalent to minimizing the clutter variance. The use of  $H(q,t)$  to minimize the number of pixel differences is analogous to the procedure employed by Watanabe (Watanabe, 1965) in showing that the Karhunen-Loeve expansion of the scene density minimizes the entropy of the squared transform coefficients over the ensemble of possible orthogonal coordinate systems. The analogous result is that the Karhunen-Loeve coordinate system minimizes the number of terms required to represent the image density.

The next chapter uses  $H(q,t)$  to tie together the effects of increasing the number of information functions, varying the scene/target ratio and increasing the number of target templates. In this work the triangle equality  $H(q,t) = H(q,p) - H(t,p)$  is used to indirectly evaluate this template to scene metric. To use the triangle equality it has been assumed that the scene density is well represented with a finite number of information functions. The final step in evaluating the detection algorithm then relates the cross-

entropy metric  $H(q,t)$  to the expected probability of error.

## Chapter VI. Processing Results

The performance factors identified in the last chapter that collectively determine the target detecting ability of the minimum cross-entropy detection algorithm will be explored in this chapter. The basis of this work is the template to scene distance  $H(q,t)$  and the triangle equality

$$H(q,t) = H(q,p) - H(t,p)$$

The performance factors will all be related to the cross-entropy  $H(q,t)$  and then using a set of test pictures the relationship between cross-entropy and probability of error is estimated. The results presented here then tie together the factors determining the error probability and allow a user to select an operating point and then project a probable performance or conversely select a required error probability and know the constraints imposed on the detection algorithm.

### Performance Factors

The number of information functions used in the minimum cross-entropy templates determines  $H(t,p)$  and by using the triangle equality also  $H(q,t)$ . Any scene can be represented exactly with a density of the form  $\exp\left(-\sum_{k=0}^{\infty} \lambda_k f_k(x,y)\right)$  where  $\{f_k(x,y)\}$  defines a complete orthogonal set of functions in  $R^2$ . Thus as information functions are added to the template

density the prior to template cross-entropy  $H(t,p)$  will increase resulting in the template to scene distance  $H(q,t)$  decreasing and eventually approaching zero. This behavior has been verified using the block tank shown in Figure 6.4 and the basic definition of cross-entropy. The block tank has a relatively high cross-entropy value of  $H(q,p) = 1.965$  due to its structure or confinement of scene energy and abrupt changes in density. Also the block tank density has no camera noise superimposed on the scene that blurs the picture and reduces template entropy. The minimum cross-entropy approximation to the block tank was computed using second through twelfth order moment information and the resulting template cross-entropy  $H(t,p)$  computed. Figure 6.1 shows the template to scene distance as a function of the number of information functions used in the template. Note that the resulting data points can be approximated with a straight line.

The minimum cross-entropy template densities produced for second through twelfth order moment information are given in Figures 6.5 through 6.15 respectively. The template densities are shown to have increasing structure and cross-entropy  $H(t,p)$  values as the number of moment constraints is increased. The block tank was used to clearly show the impact on the resulting template of the minimum cross-entropy requirement and the conflicting requirement to conform to the moment information.



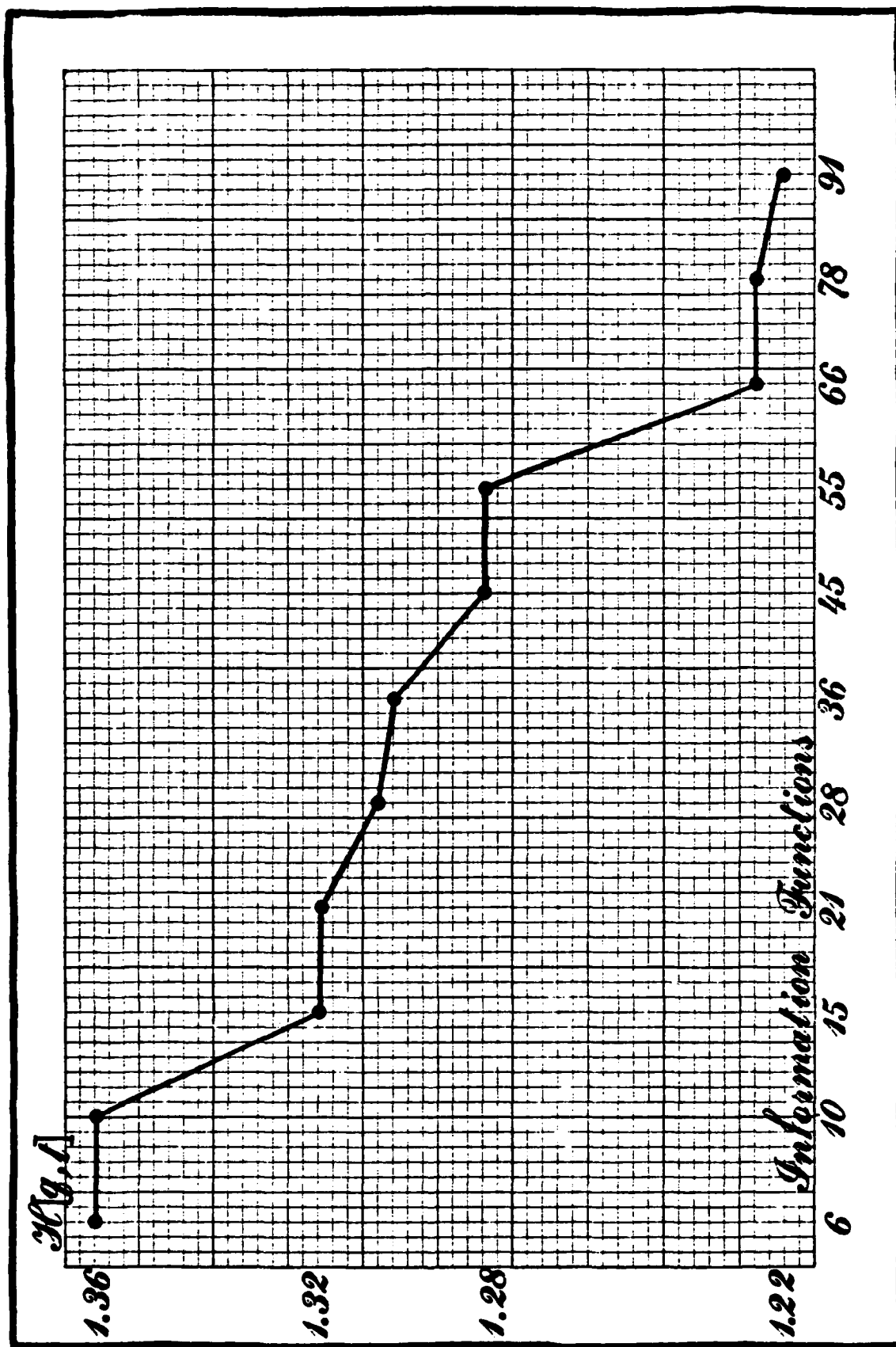


Fig. 6.1. Template to Scene Cross-Entropy versus Approximation Order

The scene cross-entropy  $H(q,p)$  measured using a uniform prior density has a finite maximum value over the unit square. Using the clutter background that produces this maximum value for scene cross-entropy allows a worst case evaluation of the relationship between  $H(q,p)$  and the scene/target ratio. The maximum value of scene cross-entropy was experimentally found to be  $H(q,p) = 2.418$  over the unit square. With the maximum cross-entropy clutter uniformly distributed over the unit square a tank was placed in the center of this cluttered scene. The size of the tank was then steadily decreased to produce Figure 6.2 which shows the increase in scene cross-entropy as the relative target size is decreased. The general shape will always hold true with the initial cross-entropy value a function of the selected target and  $H(q,p)$  asymptotically approaching the maximum value for large scene/target ratios. Note that the operating point used for the test set of images is shown on the graph.

Scene cross-entropy  $H(q,p)$  is one of two components of  $H(q,t)$  as shown by the triangle quality. This first component of the template to scene distance then defines a distance the minimum cross-entropy template density must strive for to minimize the resulting template to scene distance  $H(q,t)$ . Two factors combine to determine the template cross-entropy. The first factor is camera noise and it tends to reduce  $H(t,p)$  and thus increase the resulting template to scene distance. The second factor is the amount of informa-

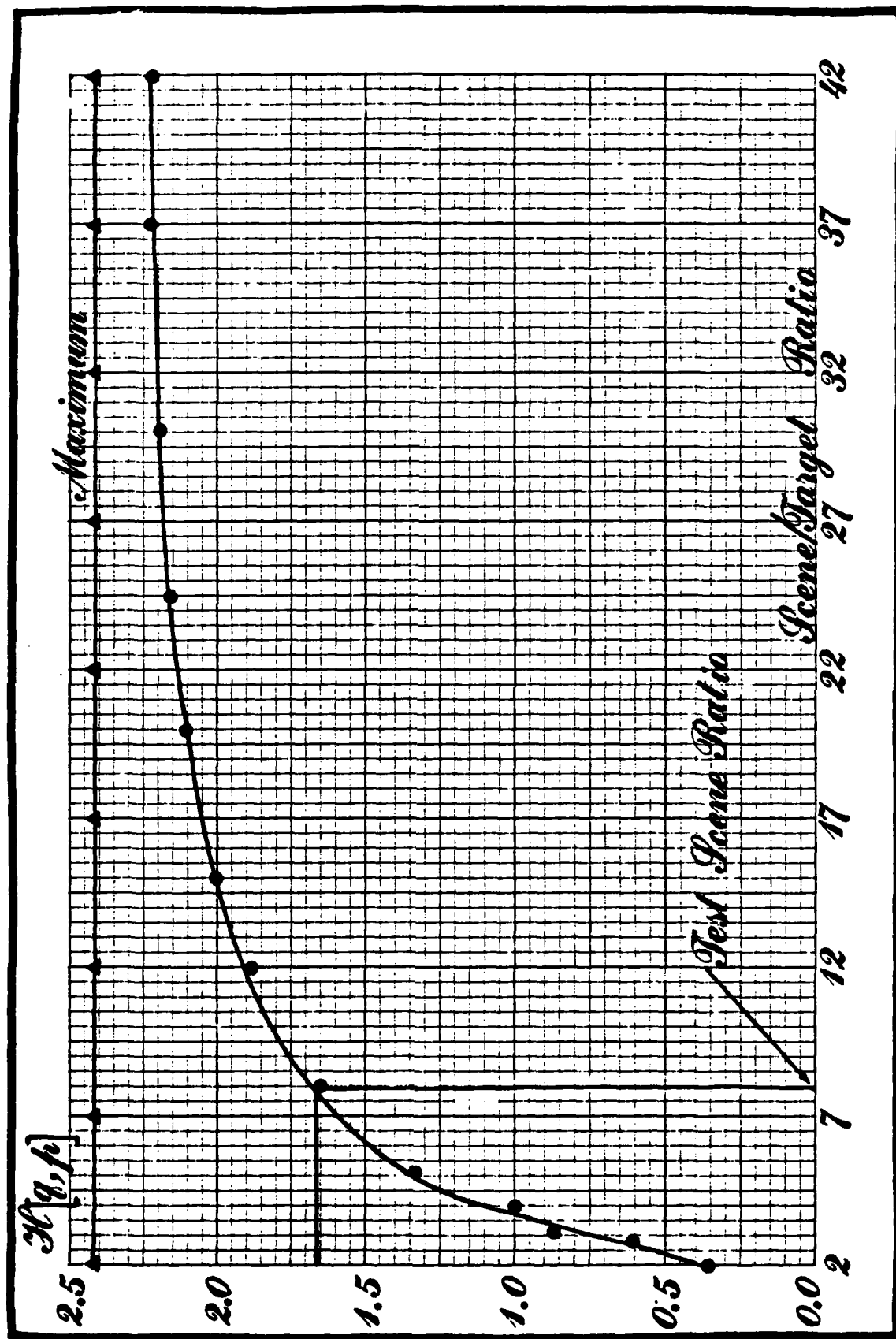


Fig. 6.2. Scene Cross-Entropy versus Scene/Target Ratio

tion used to produce the template or the number of information functions. The template cross-entropy  $H(t,p)$  moves toward  $H(q,p)$  as more information functions are added to the minimum cross-entropy template. The two concepts of relative target size and the number of information functions used in the templates can be combined into one multilevel plot as given in Figure 6.3. The straight line approximation for  $H(t,p)$  has been used to simplify the presentation. The first two performance factors, scene/target ratio and the number of information functions in the template, have been tied together with the template to scene distance  $H(q,t)$ . The next step is to estimate the relationship between  $H(q,t)$  and the probability of error.

#### Approximate Error Probability

The first step in estimating  $P(\epsilon)$  for the minimum cross-entropy detection algorithm is to evaluate performance without the interfering clutter background. Half the test scenes will contain a target and the other have only a uniform background. The target in the first half of the test scenes was placed at various locations and orientations within the information cell to insure that each target scene is unique. The initial detection rule test has thus reduced to using moments as an object descriptor an area where they have been applied extensively. Starting with early work in character recognition (Hu, 1963), (Alt, 1962) invariant moments have proven useful in locating known objects on a uniform

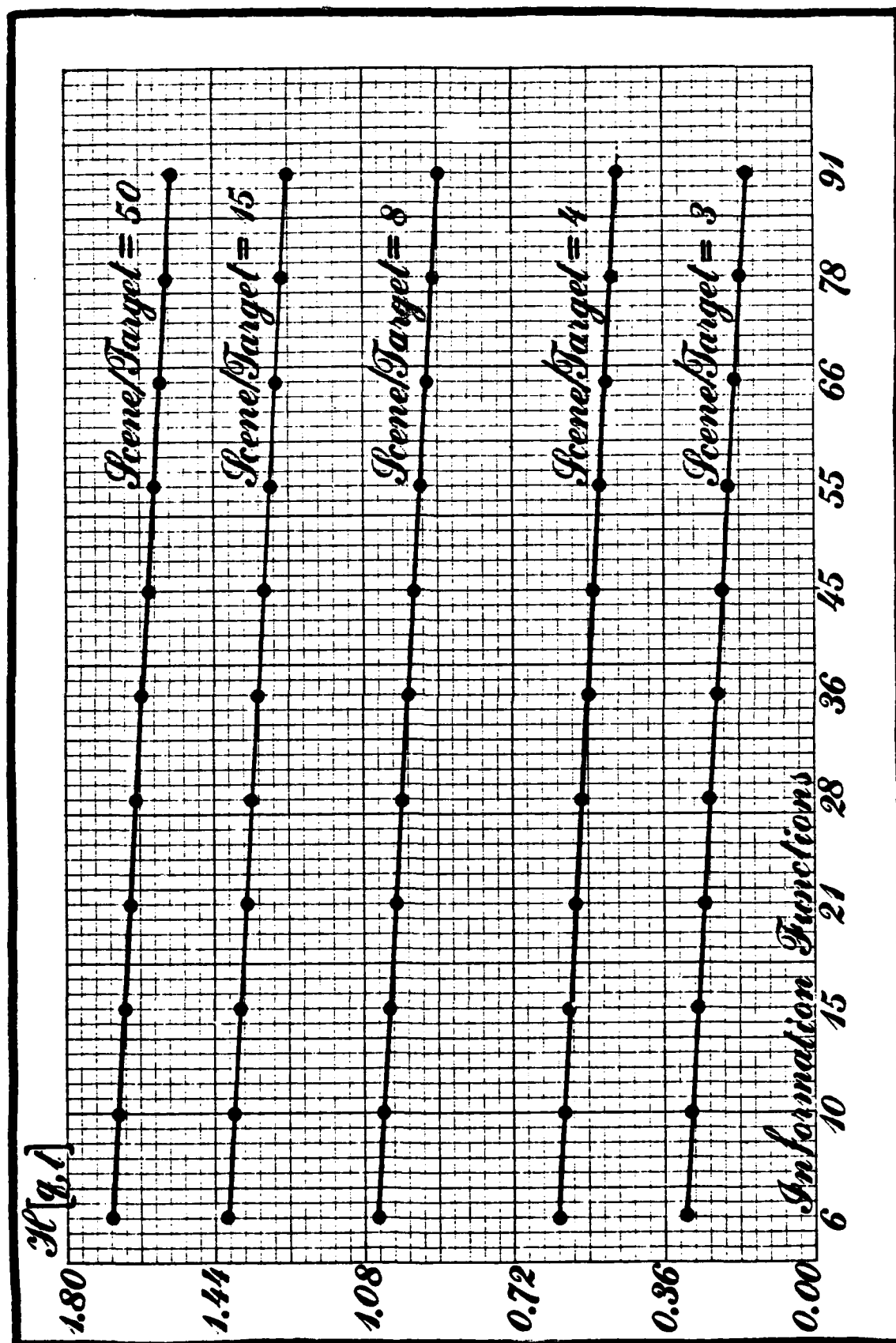


Fig. 6.3. Approximate Template to Scene Cross-Entropy

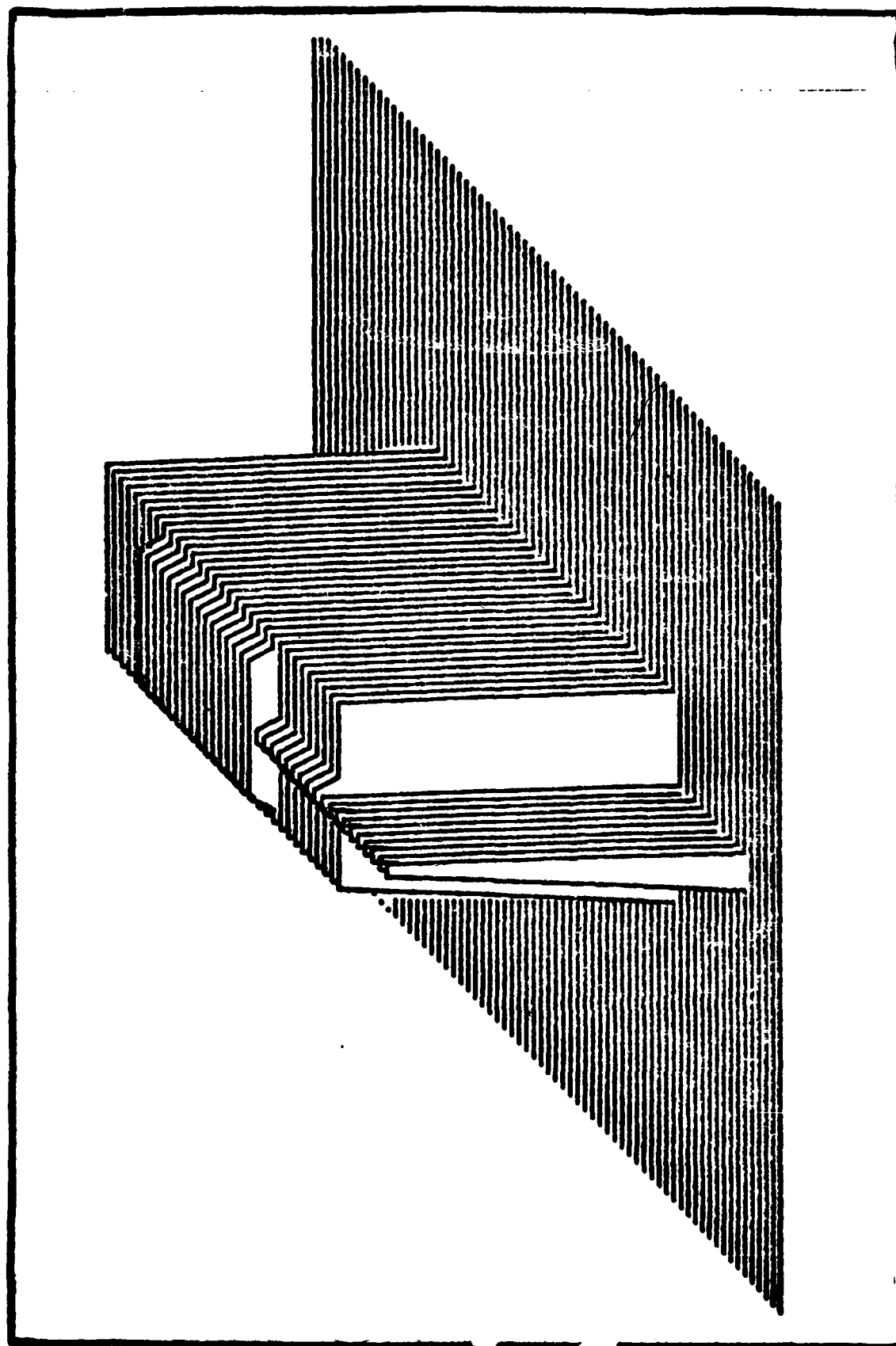


Fig. 6.4. Block Tank Density

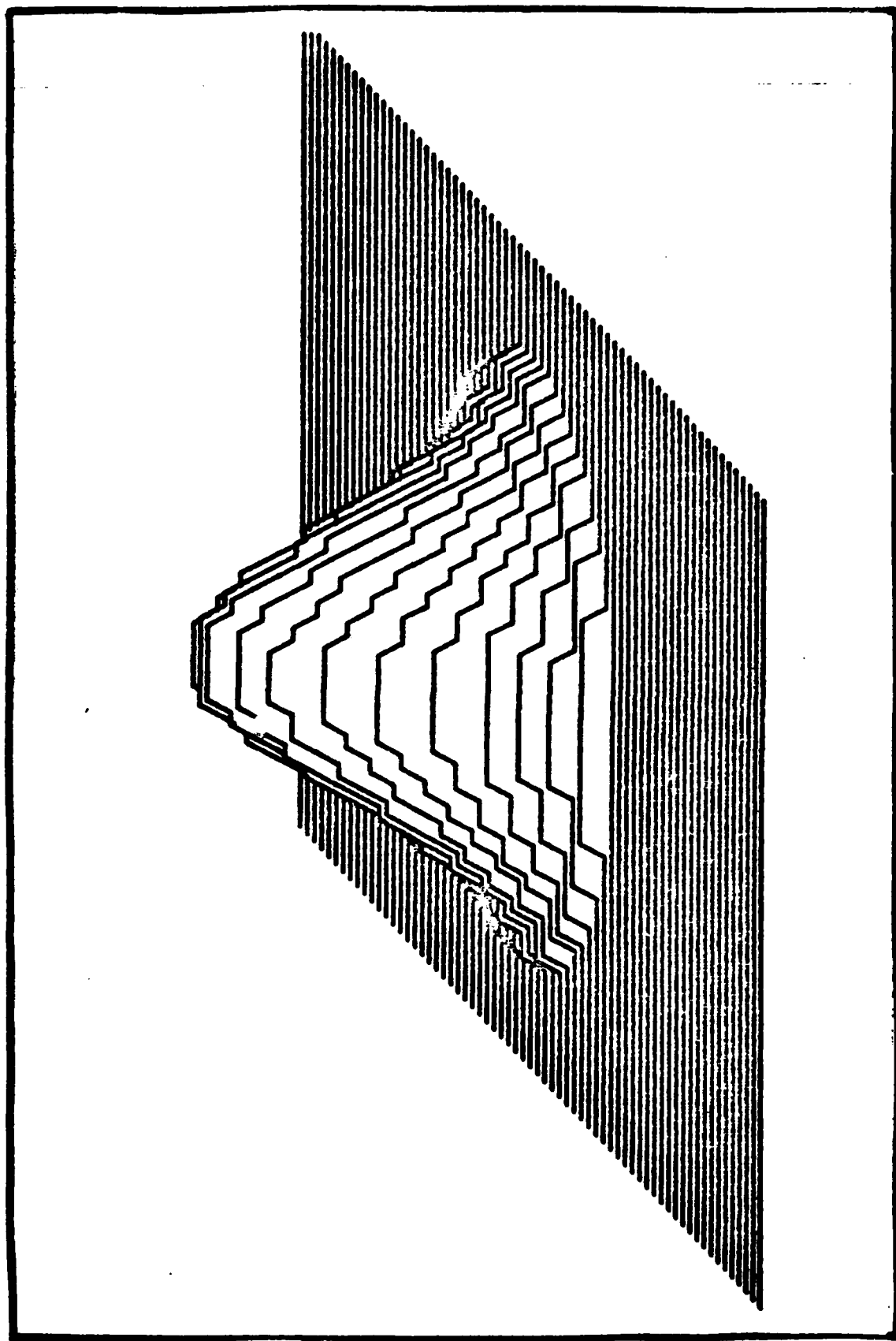


Fig. 6.5. Minimum Cross-Entropy Tank Reconstruction (6 Information Functions)

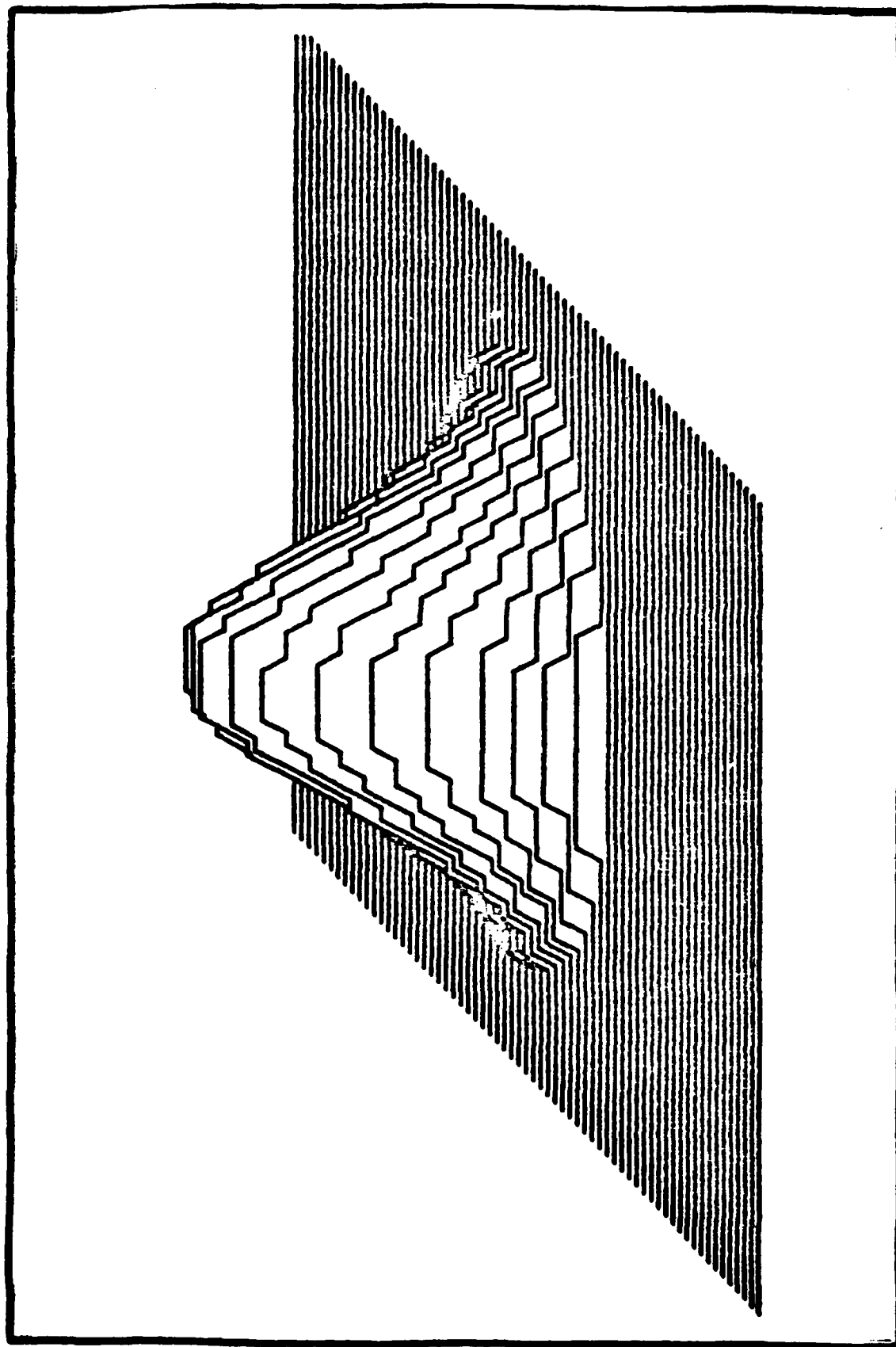


Fig. 6.6. Minimum Cross-Entropy Tank Reconstruction (10 Information Functions)



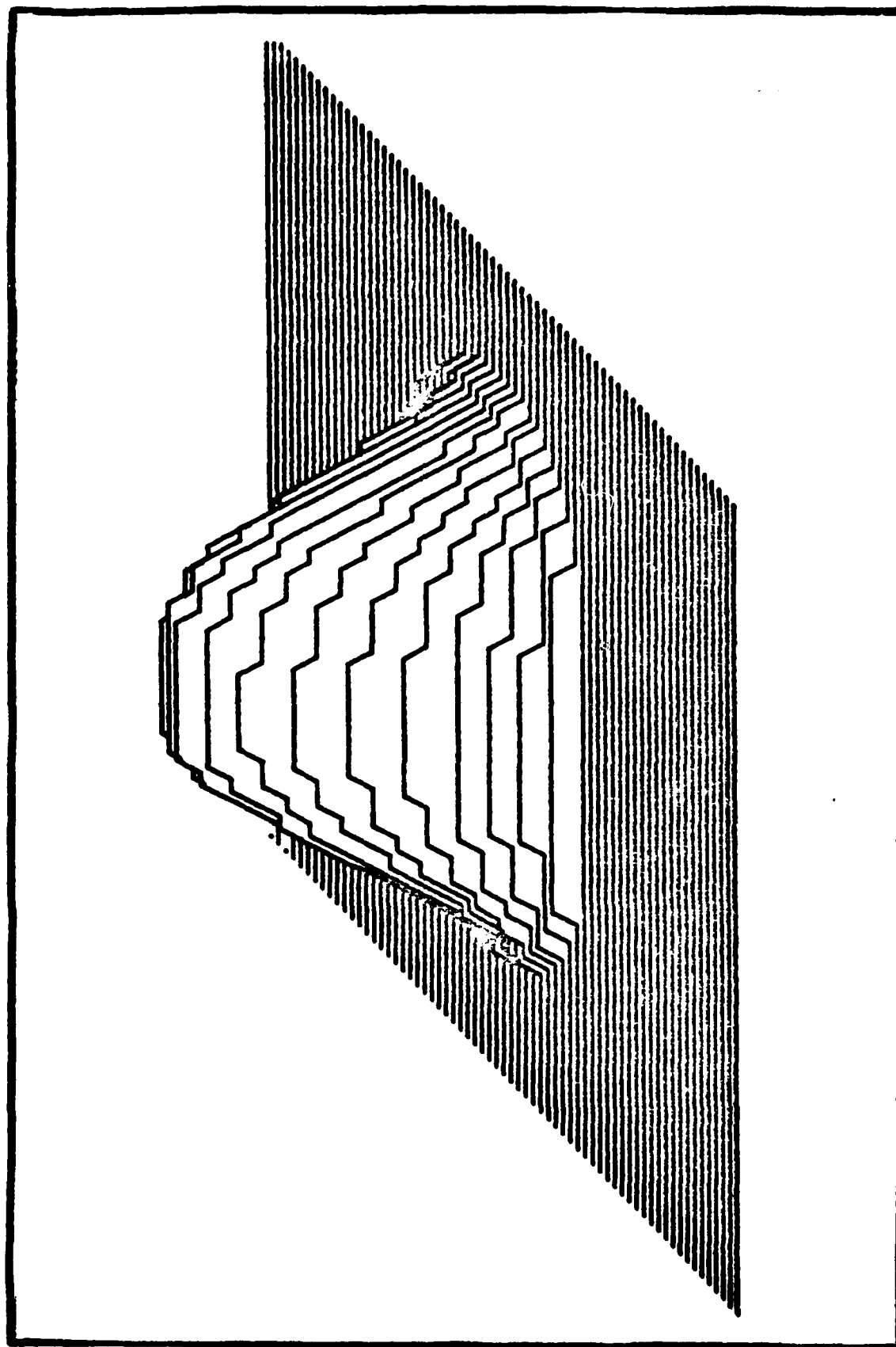


Fig. 6.7. Minimum Cross-Entropy Tank Reconstruction (15 Information Functions)

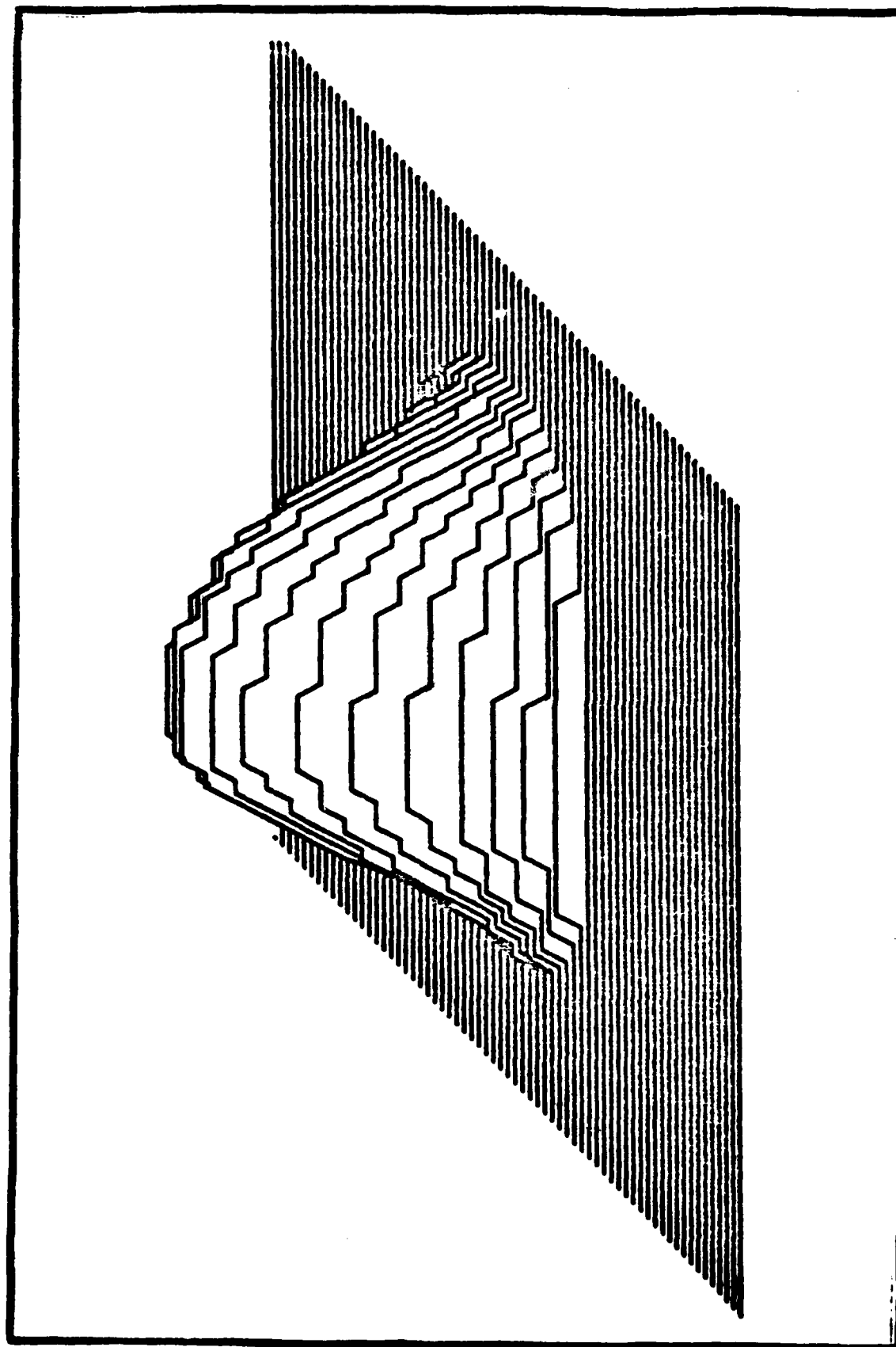


Fig. 6.8. Minimum Cross-Entropy Tank Reconstruction (21 Information Functions)

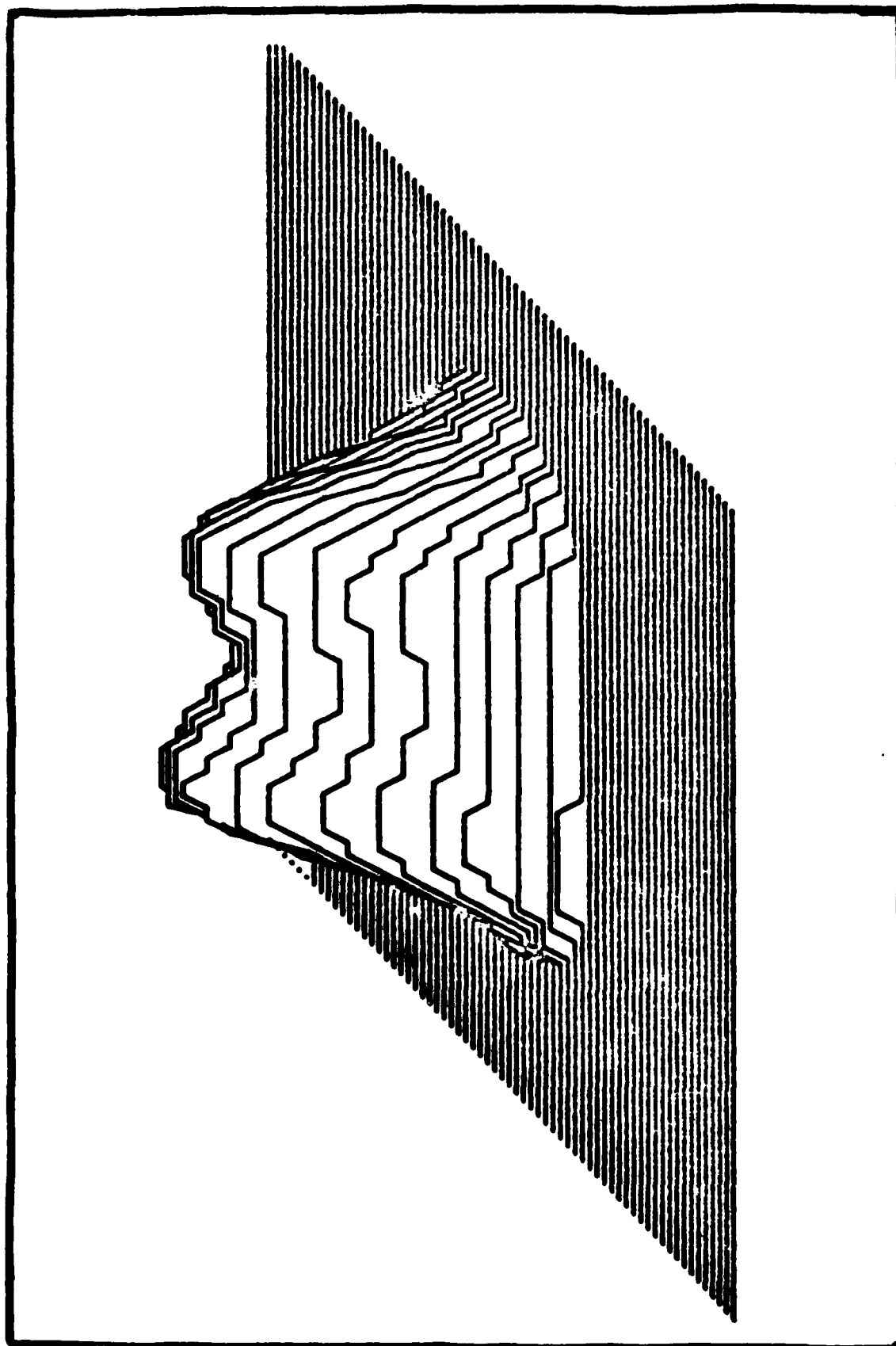


Fig. 6.9. Minimum Cross-Entropy Tank Reconstruction (28 Information Functions)

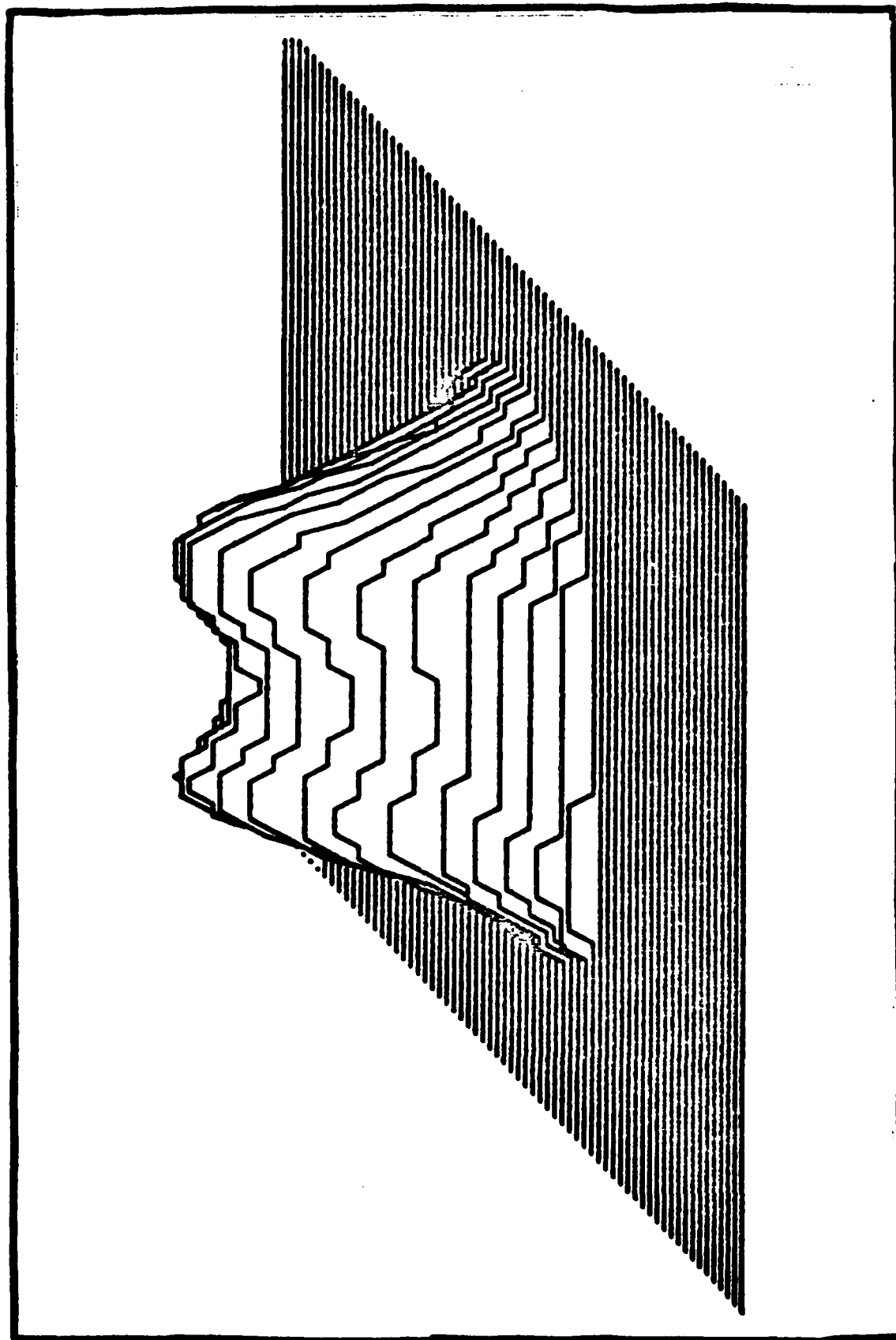


Fig. 6.10. Minimum Cross-Entropy Tank Reconstruction (36 Information Functions)

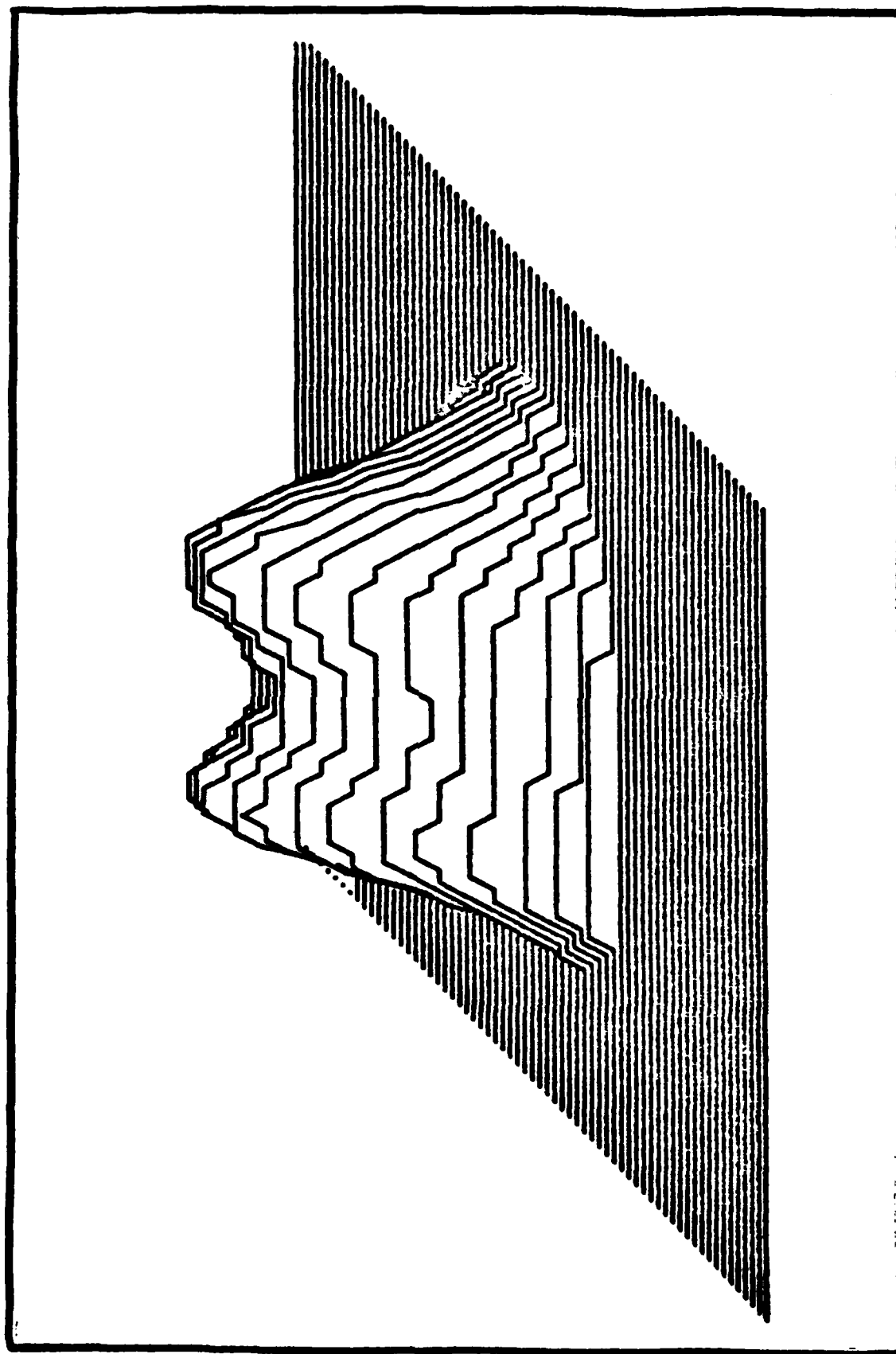


Fig. 6.11. Minimum Cross-Entropy Tank Reconstruction (45 Information Functions)

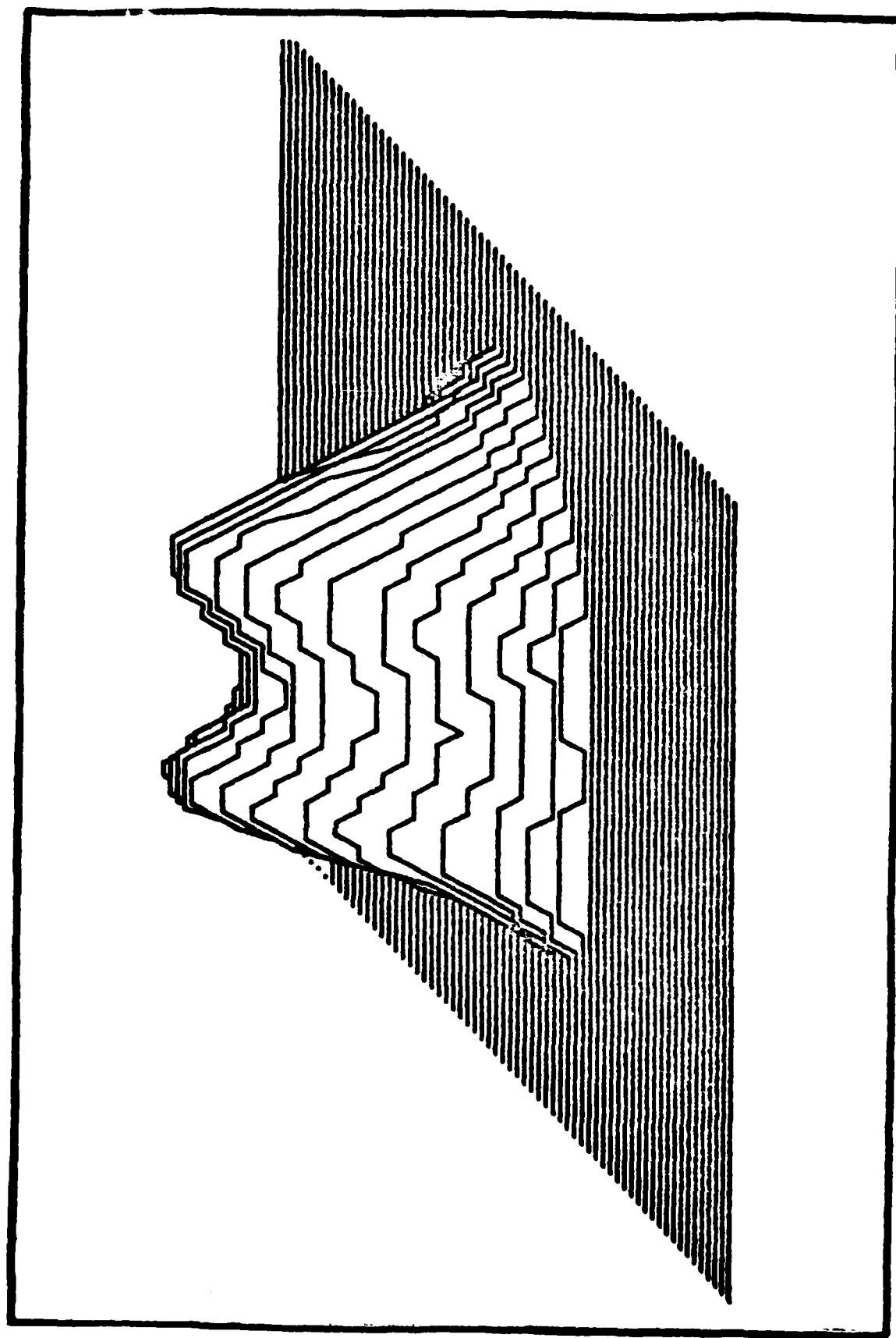


Fig. 6.12. Minimum Cross-Entropy Tank Reconstruction (55 Information Functions)

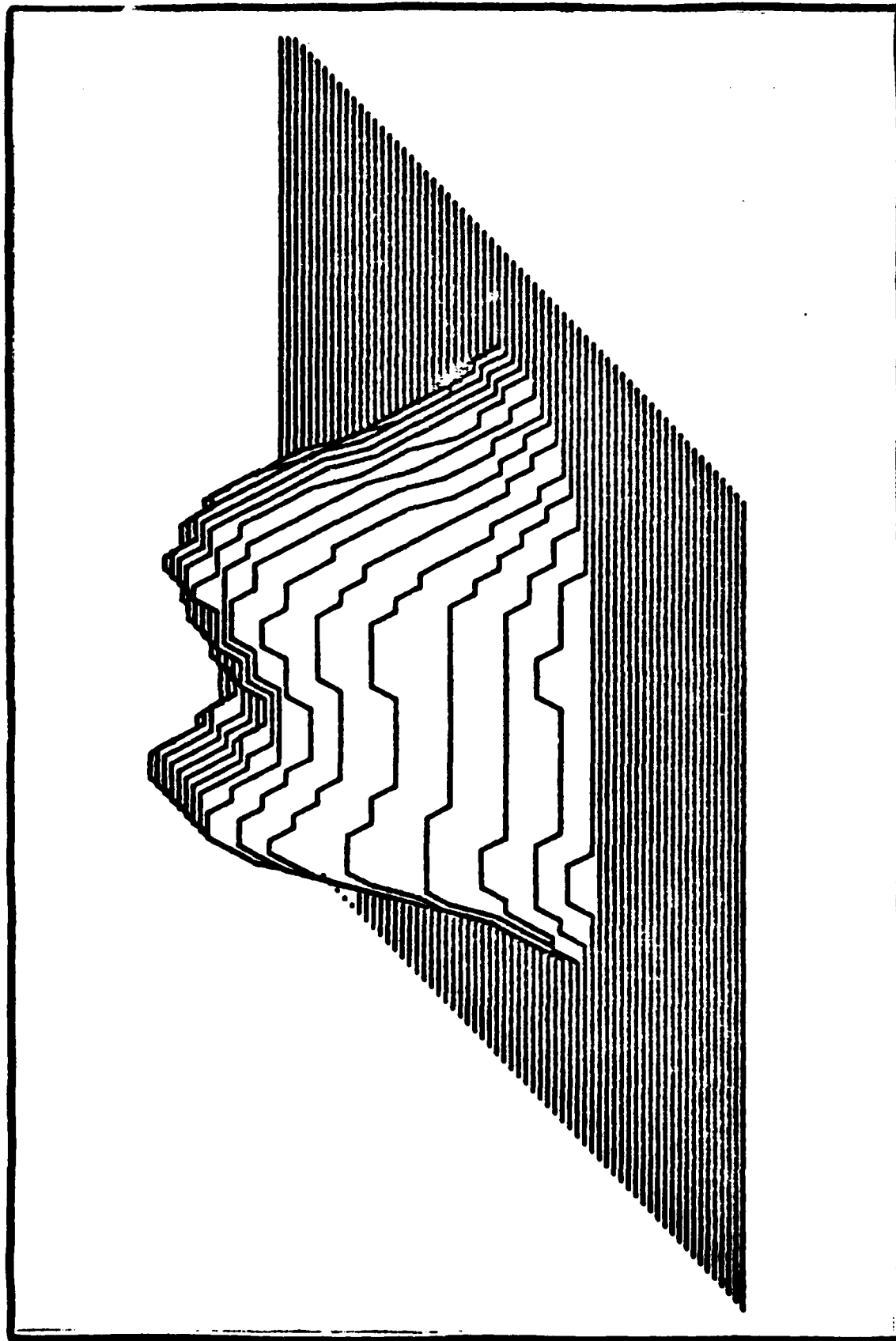


Fig. 6.13. Minimum Cross-Entropy Tank Reconstruction (66 Information Functions)

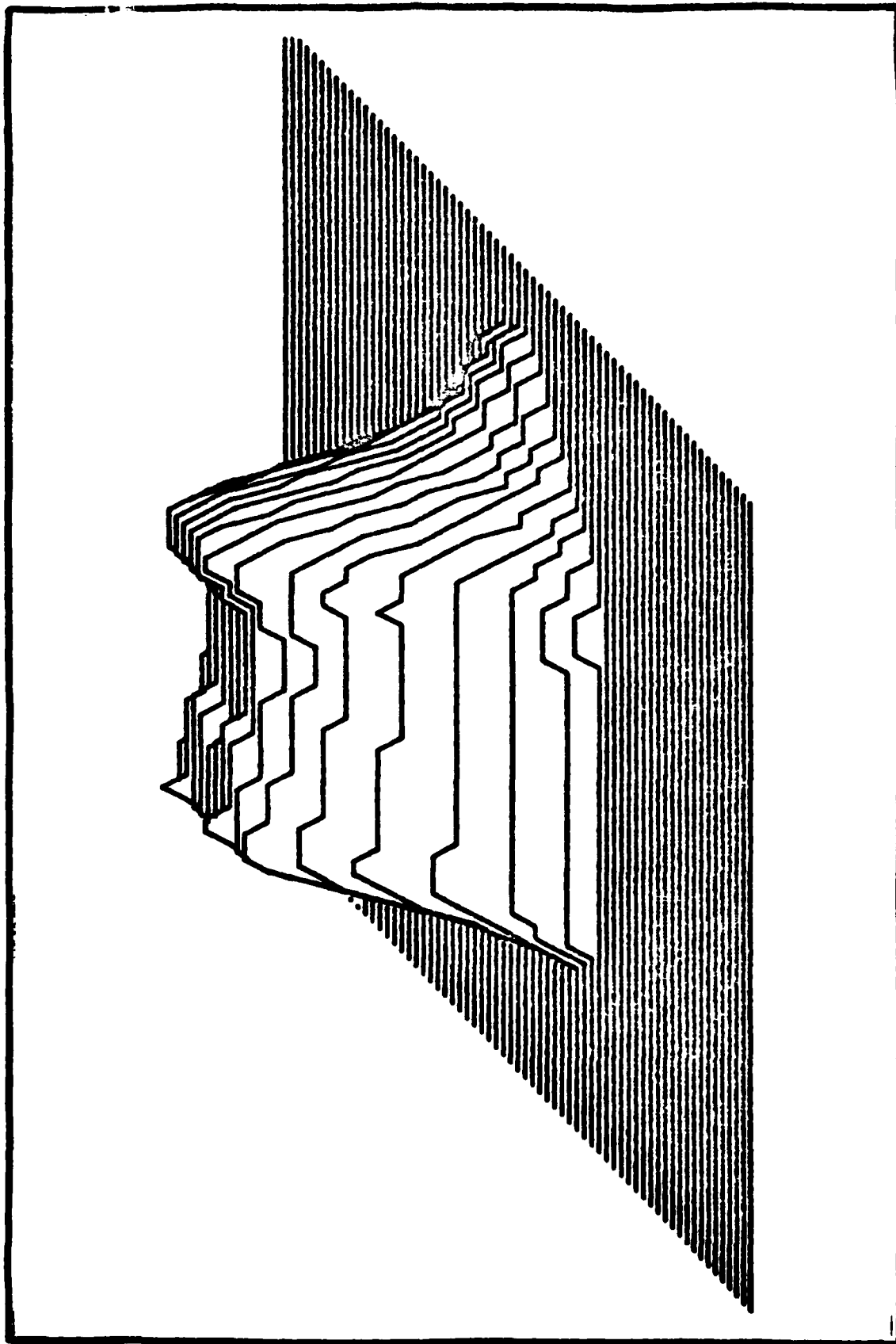


Fig. 6.14. Minimum Cross-Entropy Tank Reconstruction (78 Information Functions)



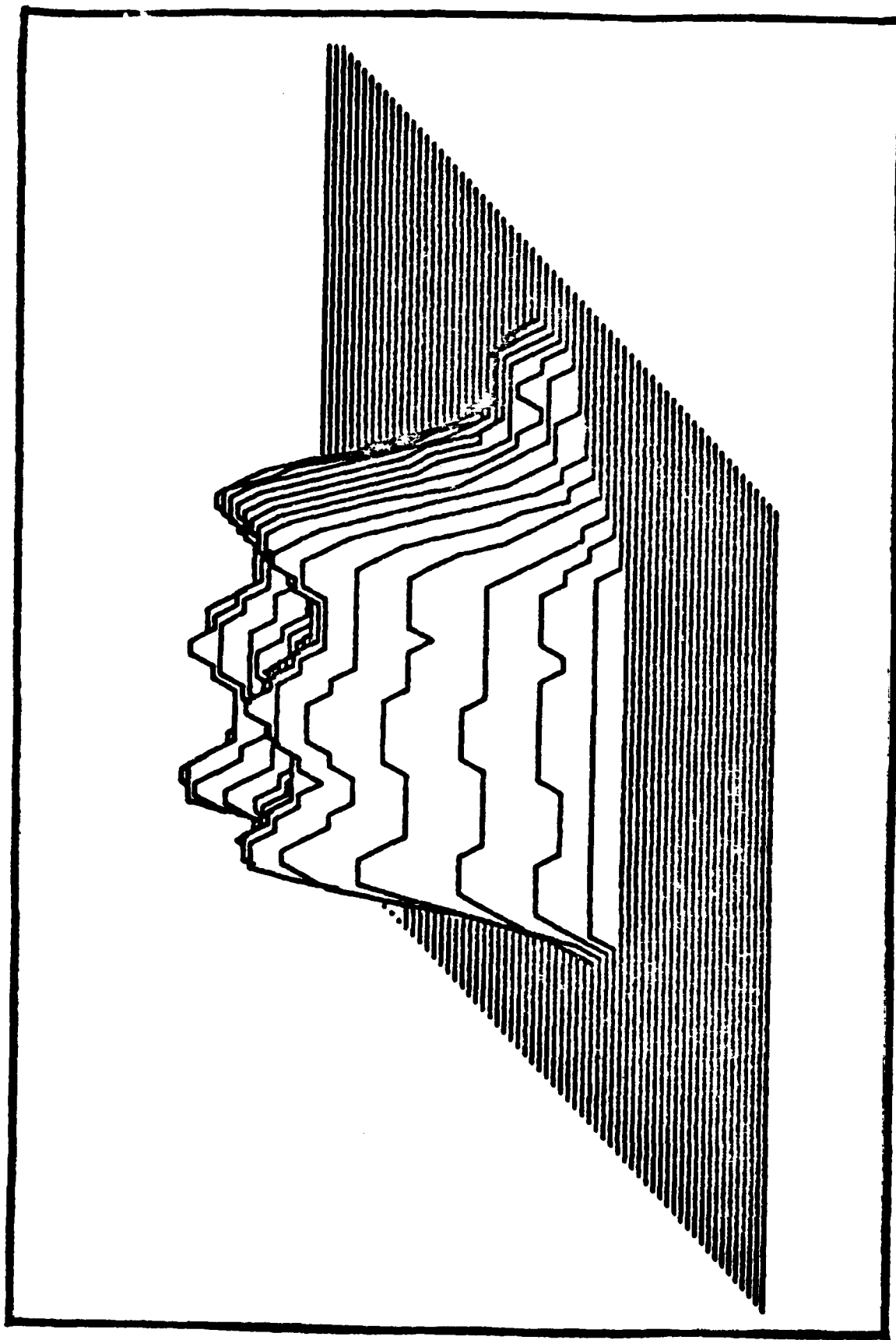


Fig. 6.15. Minimum Cross-Entropy Tank Reconstruction (91 Information Functions)

background. Automatic interpretation of ship photographs using spatial moments (Smith, 1971) has obtained performance comparable to human photointerpreters, however, again on a uniform background.

S. B. Dudani extended the moment invariant concept to the identification of three-dimensional objects in his masters thesis (Dudani, 1971) and in his doctoral dissertation (Dudani, 1973) where he conducted an experimental study of aircraft identification using moment methods. Dudani oriented his work toward video imagery and only used the information contained in the second and third order moments calculated over the image silhouette and boundary to provide the information for his target classification rule. Then he used a test set of approximately 100 images and various classifiers (Bayes, K-nearest neighbor and sequential) to show that the classifiers performance was superior to that obtained with human test subjects.

With this historical background of moments used as an object descriptor it is not surprising that the minimum cross-entropy detection algorithm performs well without an interfering clutter background. In fact, the classifier correctly recognized every scene in a test group of twenty pictures. Half of these pictures contained tanks on a uniform background and the other half only the uniform background. The minimum cross-entropy rule was shown to function correctly as an object descriptor and has performed at least as well as earlier

target detection rules on this limited set of test data.

Looking again at the history of moment methods, the next logical step would be for someone to find objects in cluttered scenes using moments. Wong and Hall (Wong, 1978) (Hall, 1979) have tried this concept by using scene invariant moments as a similarity measure in matching or registration of radar and optical images. Most researchers have, however, attempted to isolate a candidate pattern from its background by preprocessing the picture before attempting target classification. This approach to clutter occurred as stated by Nill (Nill, 1981) since it was assumed that otherwise there would be little chance of recognizing a pattern when the moments consist of contributions from the pattern and background clutter combined. The preprocessing, however, produces its own errors and destroys information in the original scene. The minimum cross-entropy detection rule provides an alternative to the preprocessing requirement by accounting for clutter with the templates and then being robust to clutter perturbations that occur in the actual scene.

The exact relationship between error probability and template to scene cross-entropy can not be established analytically and therefore must be estimated experimentally. A set of test scenes and a set of templates are both required for this experiment. The set of test scenes is represented by Figure 6.16 which shows one of the fifty tank in clutter pictures and Figure 6.17 which shows one of the fifty clutter

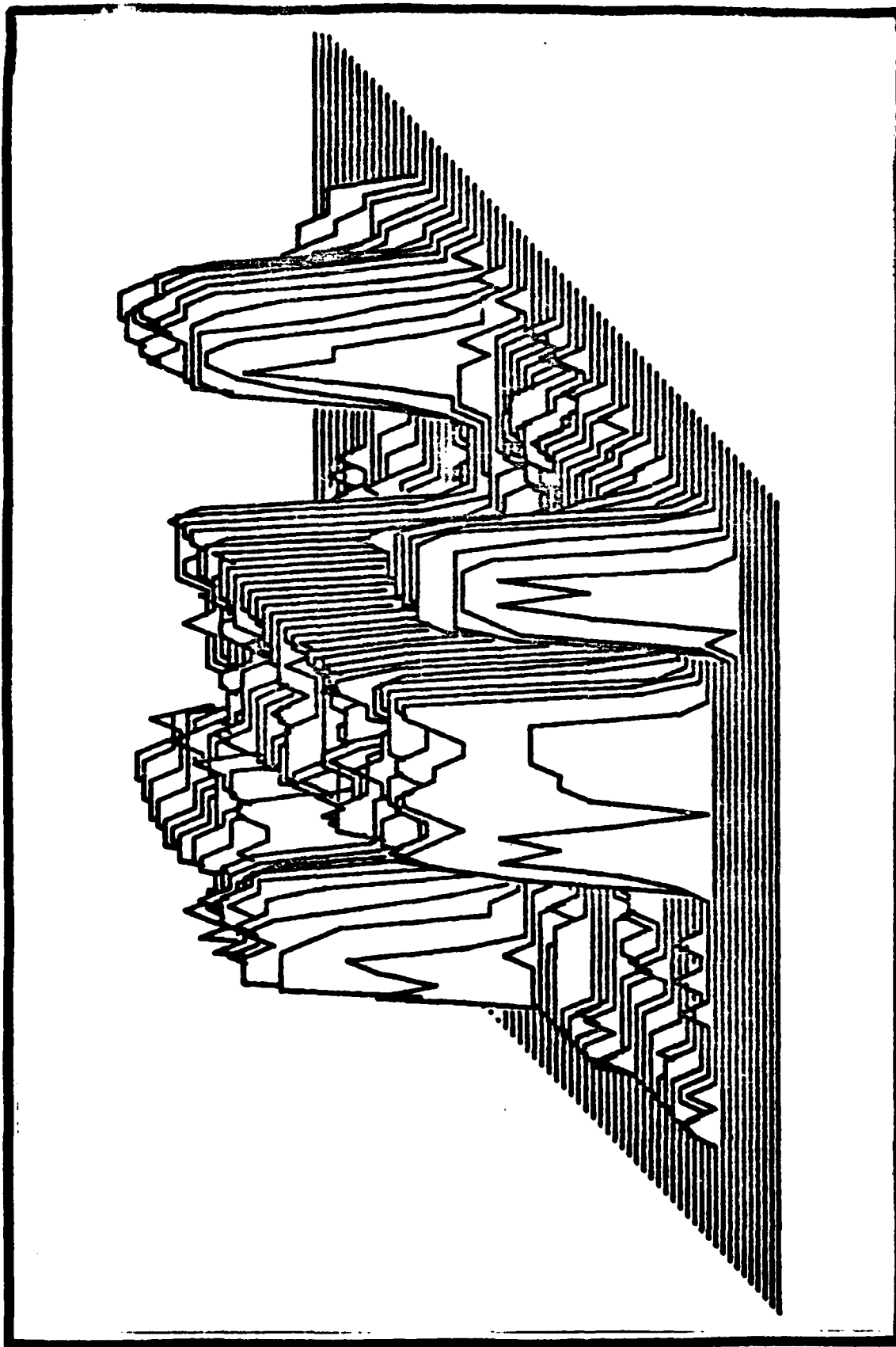


Fig. 6.16. Test Scene with Tank and Clutter

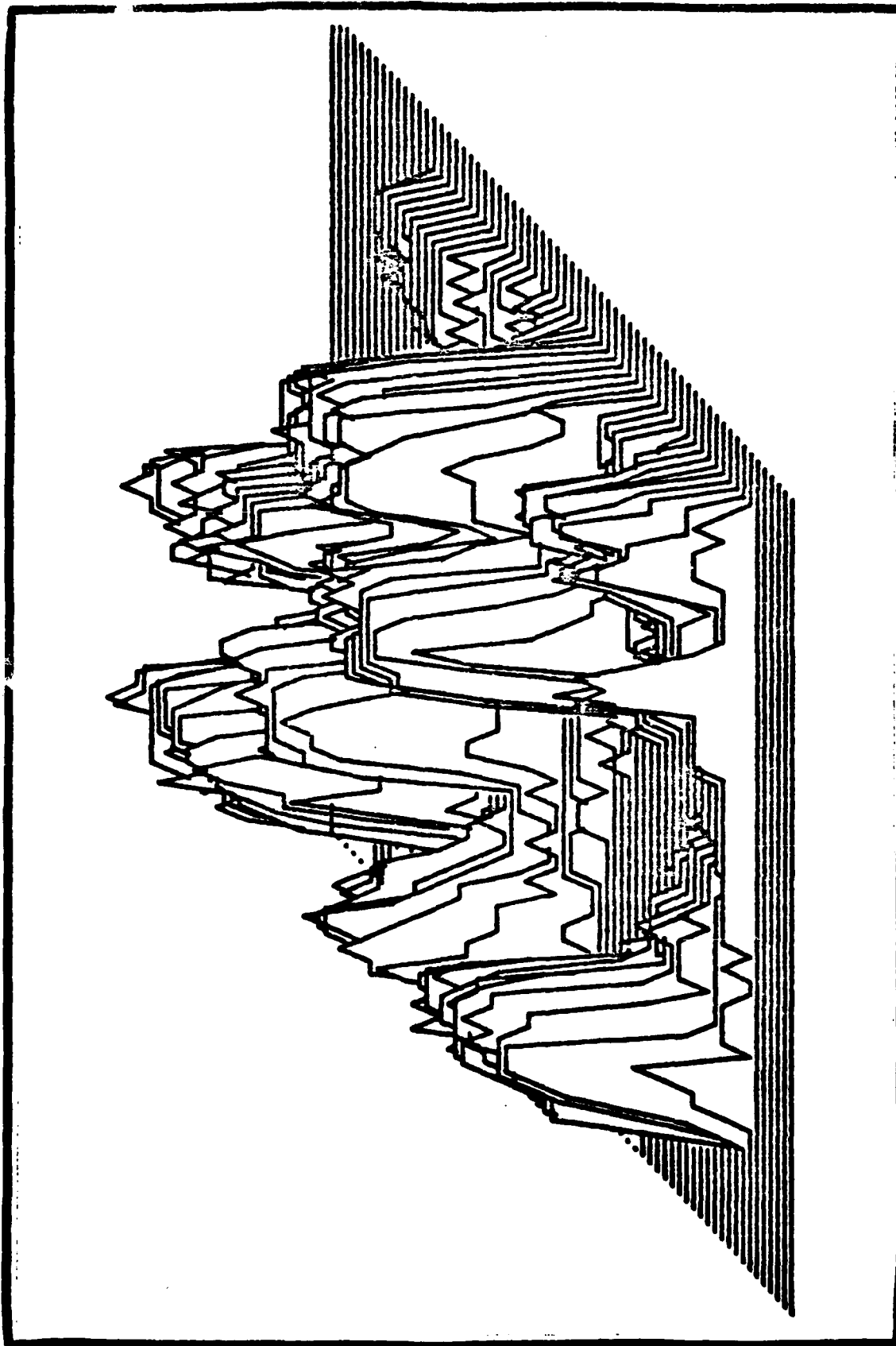


Fig. 6.i7. Test Scene with Clutter

pictures. The 100 test scenes were all unique and contained a broad spectrum of possible clutter configurations. The selected template densities are provided in Appendix A with their cross-entropy values. The templates use circular disks to represent the clutter and provide nine target-clutter alternatives for the minimum cross-entropy detection rule. The clutter model is a simplified version of one used in an Environmental Research Institute of Michigan (ERIM) Report (Wilkins, 1977) to provide a means of scene modeling and of generating "typical" scenes. The ERIM modeling procedure uses elliptical areas to represent a background scene and produce a pseudo-image whose spatial characteristics approximate those of the original image. This method of generating typical scenes is attractive and as discussed by Teague (Teague, 1980) when only moments up through second order are considered, all objects are completely equivalent to a constant irradiance ellipse having definite size, orientation, and eccentricity and centered at the object centroid. Besides making intuitive sense ERIM has experimentally found that the performance of sensors against the actual background and against the simulated background is essentially the same and thus the salient spatial features of the background have been preserved with the pseudo-image.

The circular disk is a degenerate ellipse with no orientation information and thus this clutter model is very similar to the pseudo-images used by ERIM. The templates with

clutter disks in Appendix A do suffer from superimposed camera noise which detracts from their effectiveness in the target detection algorithm. The noise impact is shown in Appendix B where the minimum cross-entropy densities corresponding to the templates of Appendix A are shown. The minimum cross-entropy densities that are used in the detection rule have low prior to template cross-entropy values because of the camera noise. The smaller template  $H(t,p)$  values result in larger average template to scene distances and larger probability of error figures. Despite the known impact of camera noise no method could be found that would remove it without distorting the template.

Using the 100 test scenes, each with cross-entropies  $H_i(q,p)$  and the 18 training templates each with minimum cross-entropies  $H_k(t,p)$  a test run of the detection rule was conducted. In the test procedure  $P(\text{tank}) = P(\text{clutter}) = \frac{1}{2}$  and the algorithm selects a template from the training set as the nearest match to each of the test scenes. Thus for each of the 100 test pictures we have the relationship

$$H_i(q,t) = H_i(q,p) - H_k(t,p)$$

$$i = 1 \dots 100$$

$$k \in \{1 \dots 18\}$$

where  $k$  is the template selected by the detection rule.

The larger the set of template densities the more likely one of the predefined template clutter configurations will correspond closely with the actual test scene and result in the correct classification of that scene. Figure 6.18 displays the performance obtained over the test set of scenes as the number of template alternatives is increased for two to eighteen. The performance improvement stops in this test of the detection algorithm but theoretically performance will continue to improve as more and more templates are available for comparison with each test scene. The departure from theory in this test can be attributed to the small size of the test scene set. Looking at Figure 5.2 relating error probability to the correlation difference/clutter standard deviation ratio shows that there is an expected slowing in performance improvement as templates are added to reduce the clutter standard deviation. The test set error probability of 0.19 corresponds to a correlation/clutter ratio of  $D/\sigma = 0.94$  in Figure 5.2 and a region of rapidly decreasing slope in the graph. Thus it is expected to require a large change in the number of templates and the resulting correlation/clutter ratio to produce further substantial improvement in the error probability for the detection rule. The improvement of  $P(\epsilon)$  with increasing numbers of templates can also be related to the improvement in the average template to scene cross-entropy of the test set defined as

$$\overline{H(q,t)} = \sum_{i=1}^{100} H_i(q,t)/100$$

111



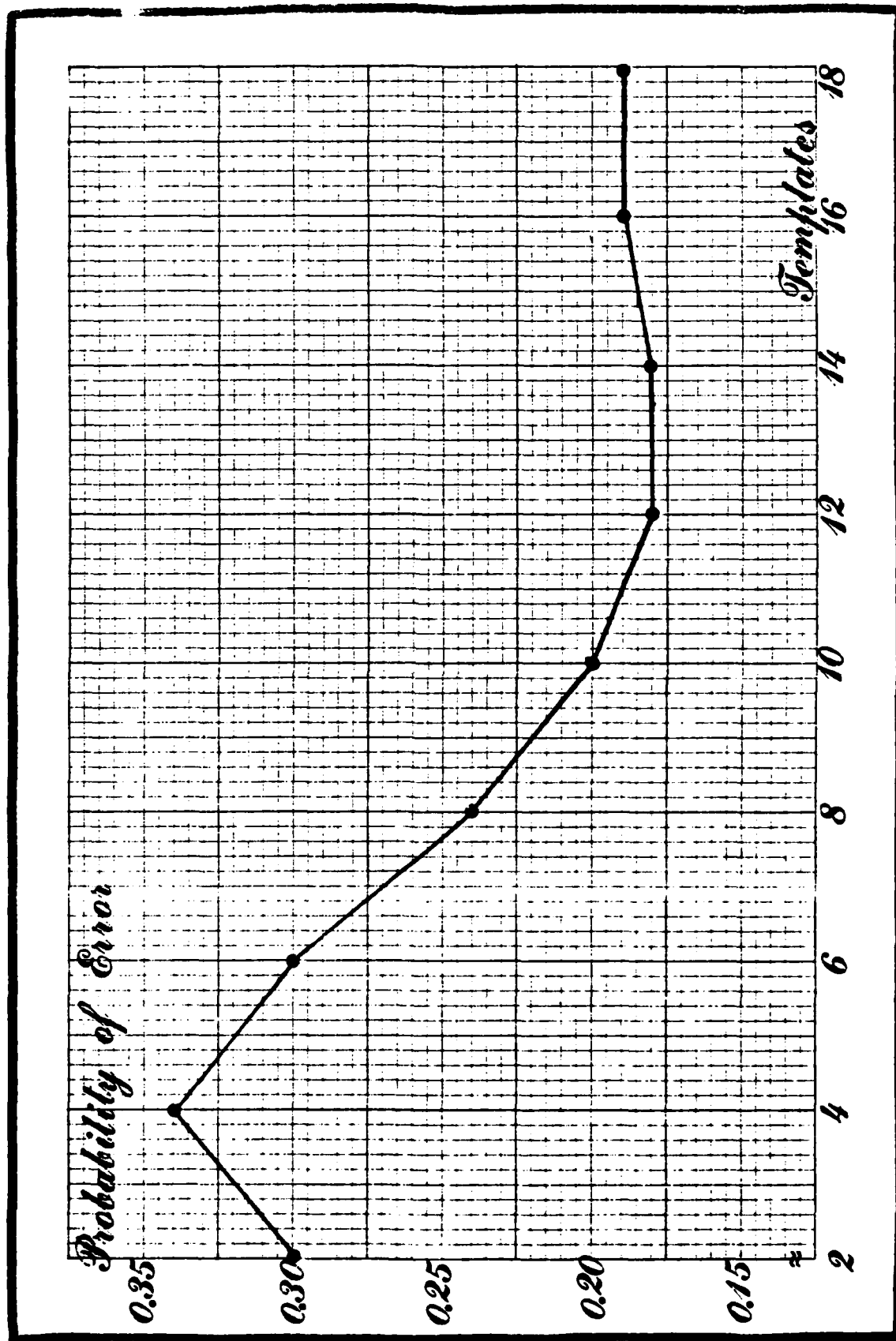


Fig. 6.18. Probability of Error versus Template Alternatives

which will decrease as better template scene matches result from the larger set of template alternatives. The template to scene cross-entropy  $H(q,t)$  thus has a well defined relationship to all three performance factors. The cross-entropy  $H(q,t)$  can also be related to the error probability which is normally used to characterize a recognition system's performance. Kovalevsky (Kovalevsky, 1980:78) explores the relationship between changes in entropy and probability of error. He was not able to find an exact functional relationship between probability of error and entropy but has established a definite relationship between these two performance indicators. The results show that for a given entropy change the error probability can vary only between definite limits and conversely for a given error probability  $P$  the entropy lies between limits that are a function of  $P$ . Since using uniform priors in the cross-entropy expressions result in an equivalence between cross-entropy and entropy, Kovalevsky's results apply to this work also since

$$H(q,t) = H(q,p) - H(t,p)$$

is exactly a change in entropy.

The bounds provided by Kovalevsky are interesting but will not allow the selection of  $H(q,t)$  based on a system error probability requirement. The partitioning of the set of test scene cross-entropy values into four equal size bins

with each having a resulting  $P(\epsilon)$  and average cross-entropy provides an approximation to the relationship between cross-entropy and probability of error. Figure 6.19 provides a broken line plot of the resulting error probability plotted against the average bin cross-entropy values. The straight line plot was obtained using the overall test set error probability as a pivot point and then providing a minimum difference compromise between the more error prone bin cross-entropy values.

Figure 6.19 clearly shows that decreasing the template to scene cross-entropy values will improve the expected error probability of the detection algorithm. Adding template alternatives and reducing the camera noise superimposed on the templates will therefore result in improved performance. Increasing the number of information functions and decreasing cell size will also result in smaller cross-entropy  $H(q,t)$  values and thus improved performance. Figures 6.19, 6.18, and 6.3 together allow a system error probability requirement to define the required performance factor values. Conversely the graphs tie together the performance factors determining error probability and given an operating point provide a good estimate of expected system performance.

The minimum cross-entropy detection rule has many attractive attributes. The rule has been shown to be optimal in a well defined information theoretic sense. Also the algorithm is computationally efficient in only computing moments and

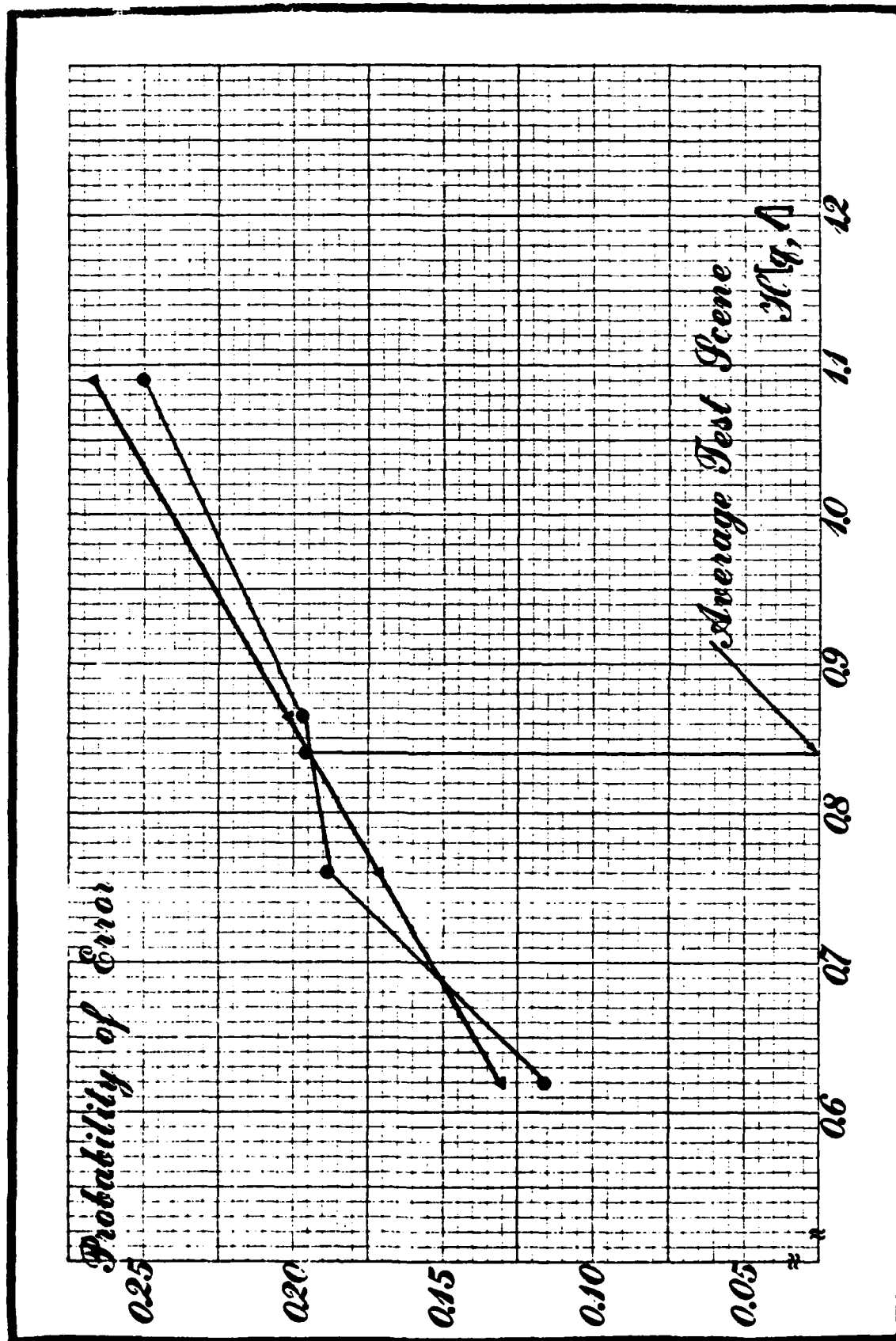


Fig. 6.19. Probability of Error versus Cross-Entropy

dot products on line. Finally as shown in this chapter the detection rule is robust in maintaining performance with a range of underlying clutter density configurations.

## Chapter VII. Summary and Future Research

### Summary

Based on the desired properties of any inference procedure stated as four consistency axioms, this dissertation has used the concept of minimum cross-entropy to develop a target in clutter detection algorithm. The algorithm uses minimum cross-entropy templates that are constructed using all available moment information, but maintaining "maximum uncertainty" with respect to unspecified information. This construction technique provides a "minimally prejudiced" template and results in a detection rule that is robust to clutter perturbations in the actual scene. The development requires information in the form of two-dimensional moments that are converted into expected values of an orthonormal set of information functions constructed with Legendre polynomials. The work is based on a constrained optimization problem and includes three procedural steps: specification of the set of template densities, solution of the constraint equations to completely define the minimum cross-entropy template and the use of cross-entropy to match actual scenes with the predefined templates.

The properties of cross-entropy minimization were reviewed showing the existence of a unique solution to the constrained optimization problem. Further, the solution density

was shown to take the following form:

$$t(x,y) = p(x,y)\exp \left\{ -\lambda_0 - \lambda_1 f_1(x,y) - \dots - \lambda_k f_k(x,y) \right\}$$

where the  $f_i(x,y)$  are information functions, the  $\lambda_i$  are the associated Lagrange multipliers, and  $p(x,y)$  is the assumed prior density. A numerical scheme based on the Cyclic Coordinate Method was presented to solve the constraint equations recast as a variational problem for a potential functional. The potential functional is concave for any trial set of Lagrange parameters and will thus always converge to a global solution.

The general target classification algorithm is developed based on the ability of cross-entropy to measure how much a scene density differs from a predefined template density. Using the triangle equality and the posterior adaptation property of minimum cross-entropy densities results in a fast on-line numerical implementation of the classification rule. The on-line processing was reduced to a dot product operation between the scene moment vector and all stored template lambda vectors followed by a search for the minimum dot product value. Conceptually, the minimum cross-entropy classifier looks for the template lambda vector most nearly orthogonal to the scene moment vector in the decision space.

Detection rule performance was examined resulting in the identification of relative target size, number of information

functions used in the template and number of template alternatives as the major performance determining factors. All three performance factors were related to a template to scene distance  $H(q,t)$  to show how these factors are interrelated in determining performance. Finally, a set of 100 test scenes was processed using the decision rule to estimate the relationship between the template to scene cross-entropy  $H(q,t)$  and the probability of error. The entire target detection procedure has been programmed and tested for computer use.

#### Future Research

The research conducted in developing the minimum cross-entropy detection algorithm surfaced several areas for continued investigation. These research areas are outlined in the following paragraphs.

The area of template selection offers large dividends in improved detection algorithm performance. A method of producing templates without camera noise will result in an immediate performance improvement. The major research area is, however, an optimum method of clutter placement within the template coupled with an analysis of the optimum number of templates for a given number of information functions and scene/target ratio. The performance impact of using ellipses rather than circles in the clutter model should also be addressed.

Using integral transforms (Wolf, 1979), the detection



problem can be moved to a transform domain which could improve error probability performance by removing some of the initial target orientation uncertainty. As examples, the magnitude of the Fourier transform of an object or function is invariant to a shift in the function and the Mellin transform is invariant to scale change in the input function. Casasent discusses these transforms and combinations such as Fourier-Mellin transforms coupled with geometrical transformations (Casasent, 1979) that provide positional, rotational, and scale invariance. In a transform domain that reduces initial target orientation uncertainty, the research would explore the optimum selection of information functions. The information function set selected would depend on the target of interest and could achieve improved probability of error performance with a simplified template model and reduced preprocessing workload.

The target detection algorithm coupled with preprocessing algorithms such as edge detectors (Abdou, 1978) is another area for further research. The information cell moments would be calculated over an image silhouette and boundary after the preprocessing operations. Using the preprocessing approach would result in a method similar to Dudani's (Dudani, 1973) that uses the preprocessing to remove unimportant information from the image and increase the template cross-entropy and thus improve error probability performance.

The research centered on characterizing two-dimensional

density functions to develop the detection rule, yet the theoretical development supports a three-dimensional characterization. Using a stereo vision system or a scanning laser rangefinder, a range image is obtained where gray level represents not brightness, but the distance from the camera to the reflecting surface in the scene (Castleman, 1979:349). The combination of a brightness image and range image produces an approximate three-dimensional image density function  $(x,y,z)$ . The three-dimensional moments of order  $p + q + s$  of the density  $(x,y,z)$  are then defined by

$$M_{pqs} = \iiint_{\text{cube}} x^p y^q z^s (x,y,z) dx dy dz$$

With the  $z$  axis perpendicular to the principal axis of the pattern, we have an immediate invariant coordinate system for the minimum cross-entropy density approximation. After defining a set of three-dimensional information functions all concepts carry forward from the two-dimensional case. In three-dimensions the solution to the constrained optimization problem takes the form

$$t(x,y,z) = p(x,y,z) \exp \left\{ -\lambda_0 - \sum_{k=1}^i \lambda_k f_k(x,y,z) \right\}$$

where  $p(x,y,z)$  is the prior density. The template clutter model extends the two-dimensional clutter ellipses to three-dimensional clutter ellipsoids and the detection rule remains unchanged. Research into a three-dimensional target

detection rule could provide a method of detecting objects in oblique aerial scenes.

Finally, this detection rule provides a means of detecting objects in cluttered scenes and we have suggested extensions that may improve the probability of error performance. Potential applications in reconnaissance, industrial robots, and imaging radar are examples that make extension of this work a viable research area.

## Bibliography

1. Abdou, Ikram E. Quantitative Methods of Edge Detection. Los Angeles, California: Image Processing Institute, University of Southern California, 1978.
2. Agmon, N., Y. Alhassid and R. D. Levine. "An Algorithm for Finding the Distribution of Maximal Entropy," Journal of Computational Physics, 30: 250-258 (1979).
3. Alt, Franz L. "Digital Pattern Recognition by Moments," Journal of the Association for Computing Machinery Vol 9: 240-258 (April 1962).
4. Bazaraa, Mokhtar S. and C. M. Shetty. Nonlinear Programming Theory and Algorithms. New York: John Wiley and Sons, 1979.
5. Burg, John P. Maximum Entropy Spectral Analysis. Ph. D. Dissertation. Department of Geophysics, Stanford University, Stanford, California, 1975.
6. Casasent, David and Demetri Psaltis. "New Optical Transforms for Pattern Recognition," Proceedings of the IEEE Vol 65: 77-84 (January 1977).
7. Castleman, Kenneth R. Digital Image Processing. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1979.
8. Courant R. and D. Hilbert. Methods of Mathematical Physics. New York: Interscience Publishers, Inc., 1953.
9. Crain, Bradford R. "Estimation of Distributions Using Orthogonal Expansions," Annals of Statistics, 2: 454-463 (May 1974).
10. -----. "An Information Theoretic Approach to Approximating a Probability Distribution," SIAM Journal of Applied Mathematics, 32: 339-346 (March 1977).
11. -----. "More on Estimation of Distributions Using Orthogonal Expansions," Journal of the American Statistical Association, 71: 741-745 (September 1976).
12. Csiszar, Imre. "I-Divergence Geometry of Probability Distributions and Minimization Problems," Annals of Probability, 3: 146-158 (February 1975).

13. Dodes, Irving A. Numerical Analysis for Computer Science. New York: Elsevier North-Holland, Inc., 1978.
14. Duda, Richard O. and Peter E. Hart. Pattern Classification and Scene Analysis. New York: John Wiley and Sons, 1973.
15. Dudani, Sahibsingh A. Moment Methods for the Identification of Three-Dimensional Objects from Optical Images. M. S. Thesis, Ohio State University, 1971.
16. -----. An Experimental Study of Moment Methods for Automatic Identification of Three-Dimensional Objects from Television Images. Ph. D. Dissertation, Ohio State University, 1973.
17. Frieden, Roy B. "Image Restoration Using a Norm of Maximum Information," Proceedings of the Society of Photo-Optical Instrumentation Engineers, 207: San Diego, California (August, 1979).
18. -----. "Restoring with Maximum Likelihood and Maximum Entropy," Journal of the Optical Society of America, 62: 511-518 (April 1972).
19. Gagnon, R. A. Detection and Identification of Objects Embedded in Cluttered Fields: A Reconnaissance Problem. Ph. D. Dissertation, Air Force Institute of Technology, 1975.
20. Gokhale, D. V. and S. Kullback. The Information in Contingency Tables. New York: Marcel/Dekher, 1978.
21. Good, I. J. "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," Annals of Mathematical Statistics, 34: 911-934 (1963).
22. Gray, Robert M. and John E. Shore. "Minimum Cross-Entropy Pattern Classification and Cluster Analysis," Naval Research Laboratory Memorandum Report 4207, (April 1980).
23. Guiasu, Silviu. Information Theory with Applications. New York: McGraw-Hill, Inc., 1977.
24. Hall, Charles F. Digital Color Image Compression in a Preceptual Space. Ph. D. Dissertation, University of Southern California, 1978.
25. Hall, Ernest L. Computer Image Processing and Recognition. New York: Academic Press, 1979.

26. Harley, T. J., L. N. Kanal and N. C. Randall. "System Considerations for Automatic Imagery Screening," Machine Recognition of Patterns, Edited by Ashok K. Agrawala: 75-91 (1977).
27. Hu, Ming-Kuei. "Visual Pattern Recognition by Moment Invariants," IRE Transactions on Information Theory, 8: 179-187 (February 1962).
28. Jaynes, Edwin T. "Information Theory and Statistical Mechanics I," Physical Review 106: 620-630 (May 1957).
29. -----. "Prior Probabilities," IEEE Transactions on Systems, Science and Cybernetics, SSC-4: 227-241 (September 1968).
30. Johnson, Rodney W. "Axiomatic Characterization of the Directed Divergences and Their Linear Combinations," IEEE Transactions on Information Theory, IT-25: 709-716 (November 1979).
31. -----. "Determining Probability Distributions by Maximum Entropy and Minimum Cross-Entropy," Proceedings APL 79: 24-29 (1979).
32. Johnson, Rodney W. and John E. Shore. "Properties of Cross-Entropy Minimization," Naval Research Laboratory Memorandum Report 4189, (April 1980).
33. Kanazawa, Tyle T. The Application of Two-Dimensional Moment Invariants to Image Signal Processing and Pattern Recognition. M. S. Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, 1980.
34. Kovalevsky, Vladimir A. Image Pattern Recognition. New York: Springer-Verlag New York Inc., 1980.
35. Kreyszig, Erwin. Introductory Functional Analysis with Applications. New York: John Wiley and Sons, 1978.
36. Kullback, Soloman. Information Theory and Statistics. New York: John Wiley and Sons, Inc., 1959.
37. Luenberger, David G. Optimization by Vector Space Methods. New York: John Wiley and Sons, Inc., 1969.
38. Markel, J. D. and A. H. Gray, Jr. Linear Prediction of Speech. New York: Springer-Verlag, 1976.
39. McEliece, Robert J. The Theory of Information and Coding. Reading, Massachusetts: Addison-Wesley Publishing Company, 1977.

40. Miller, James E. Continuous Density Approximation on a Bounded Interval Using Information Theoretic Concepts. Ph. D. Dissertation. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, 1980.
41. Minerbo, Gerald. "MENT: A Maximum Entropy Algorithm for Reconstructing a Source from Projection Data," Computer Graphics and Image Processing, 10: 48-68 (1979).
42. Neyman, J. "'Smooth' Test for Goodness of Fit," Skand. Aktuarietidsh, 20: 149-199 (1937).
43. Nill, Norman B. "Applications of Moments to Image Understanding," Techniques and Applications of Image Understanding SPIE Vol 281. Edited by James J. Pearson: 126-131 (April 1981).
44. Ortega, James M. and W. C. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables. New York: Academic Press, 1970.
45. Papoulis, Athanasios. Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill Book Company, 1965.
46. Pratt, William K. Digital Image Processing. New York: John Wiley and Sons, 1978.
47. Rey, William J. J. Robust Statistical Methods. Berlin: Springer-Verlag, 1978.
48. Saaty, T. L. and J. Bram. Nonlinear Mathematics. New York: McGraw-Hill, Inc., 1964.
49. Shannon, Claude E. "A Mathematical Theory of Communication," Bell System Technical Journal, 27: 379-423 (1948).
50. ----- "A Mathematical Theory of Communication," Bell System Technical Journal, 27: 623-656 (1948).
51. Shore, John E. "Minimum Cross-Entropy Spectral Analysis," IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-29: 230-237 (April 1981).
52. Shore, John E. and Rodney W. Johnson. "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Transactions on Information Theory, IT-26: 26-37 (January 1980).
53. Smith, Fred W. and Margaret H. Wright. "Automatic Ship Photo Interpretation by the Method of Moments," IEEE Transactions on Computers, C-20: 1089-1095 (September 1971).

54. Teague, Michael R. "Image Analysis Via the General Theory of Moments," Journal of the Optical Society of America, 70: 920-930 (August 1980).
55. Watanabe, Satoshi. "Karhunen-Loeve Expansion and Factor Analysis: Theoretical Remarks and Applications," Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes: 635-660 (1965).
56. ----- "Pattern Recognition as a Quest for Minimum Entropy," Pattern Recognition, 5: 381-387 (1981).
57. Whitmore, George D. (Editor). Manual of Photogrammetry (Third Edition). Falls Church: American Society of Photogrammetry, 1966.
58. Wilkins, Lowell. Statistical Analysis of Terrain Background Measurements Data. Final Report, Ann Arbor, Michigan: Environmental Research Institute of Michigan, 1977.
59. Wolf, Kurt B. Integral Transforms in Science and Engineering. New York: Plenum Press, 1979.
60. Wong, Robert Y. and Ernest L. Hall. "Scene Matching with Invariant Moments," Computer Graphics and Image Processing, 8: 16-24 (1978).
61. Young, David M. and Robert T. Gregory. A Survey of Numerical Mathematics Volume I. Reading, Massachusetts: Addison-Wesley Publishing Company, 1972.



### Appendix A. Template Densities

The eighteen template scenes used in Chapter Six to evaluate the approximate probability of error are shown in this appendix. Moments were taken from these template scenes to define a set of nonlinear constraint equations for each density. The constraint equations were solved for the lambda vector using the cyclic coordinate search method. The resulting lambda vector then completely defines a corresponding minimum cross-entropy template that is provided in Appendix B.

These templates were produced from photographs and are stored in the computer as 256 x 256 integer arrays. The gray scale is confined to sixteen levels in the sampling process which accounts for the abrupt changes in intensity seen in the templates. Note the camera noise superimposed on these templates and their resulting low cross-entropy values.

The odd number templates represent clutter scenes while the even number templates represent a tank with various clutter backgrounds. Each template pair (ex. 1 and 2) represents a tank-clutter alternative for the detection rule. The clutter seen in these templates comes from dark circular disks distributed in the scene that represent various possible clutter configurations.

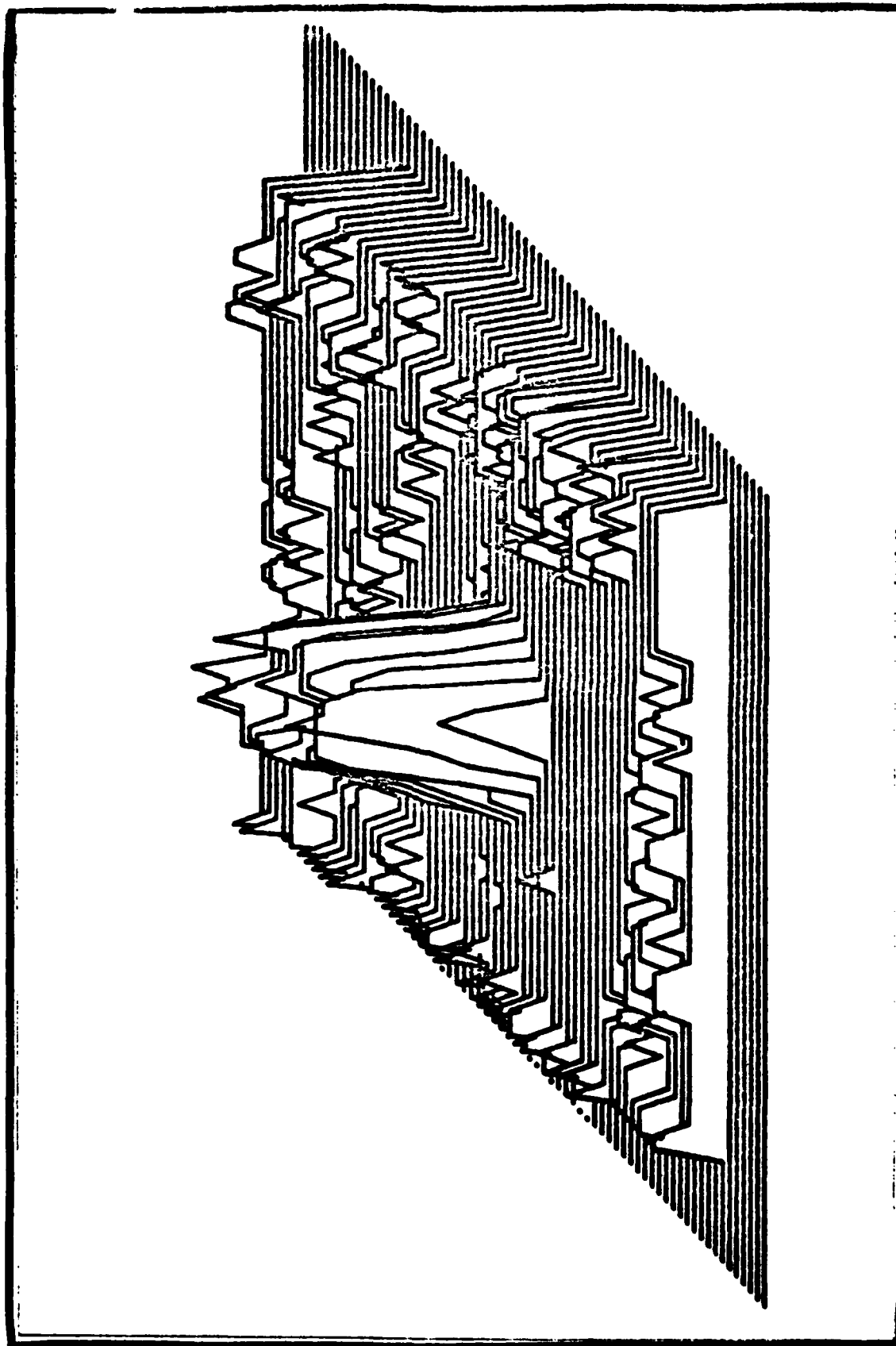


Fig. A.1. Template Density 1 with Cross-Entropy = 0.71460

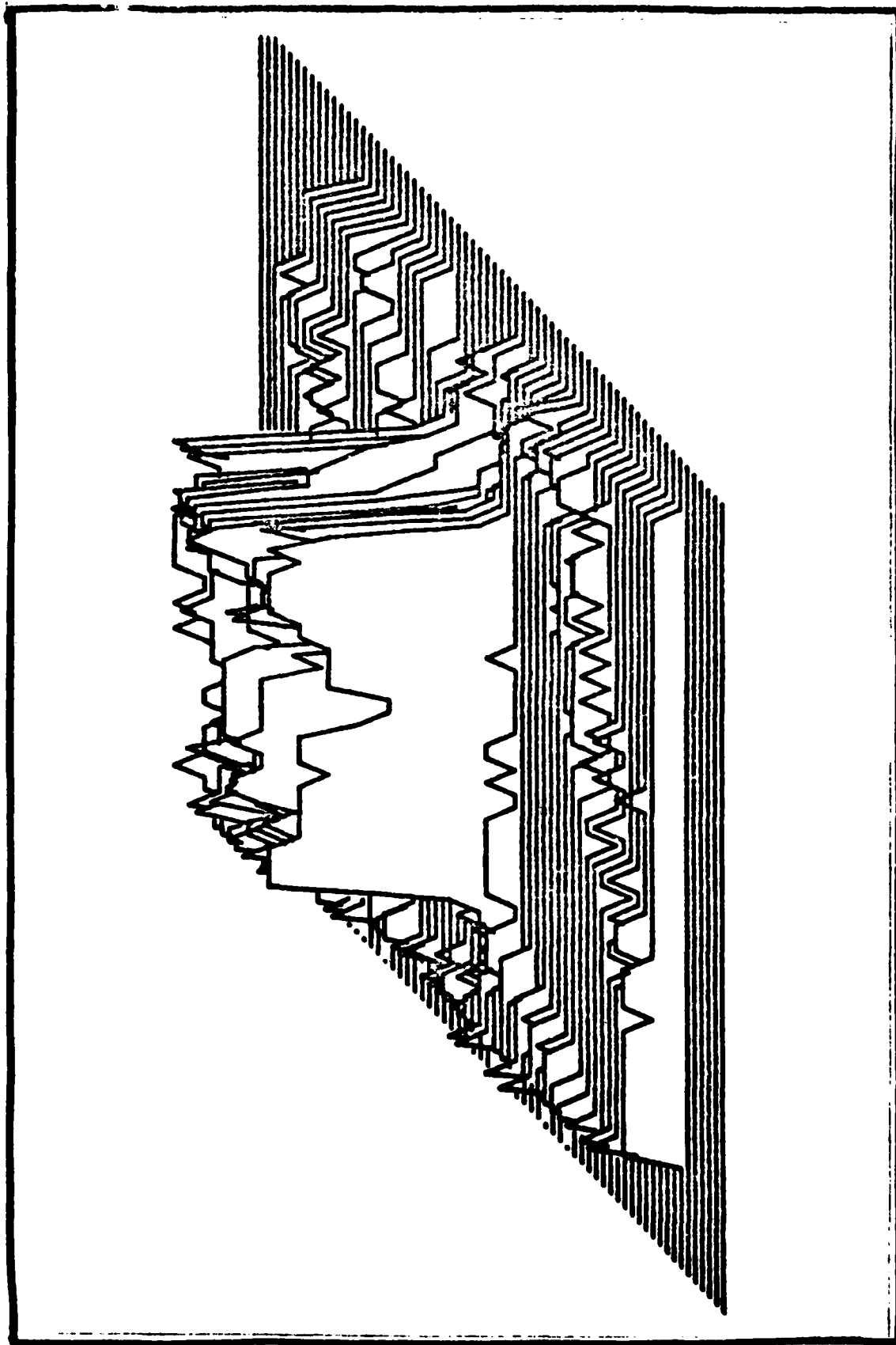


Fig. A.2. Template Density 2 with Cross-Entropy = 0.86036

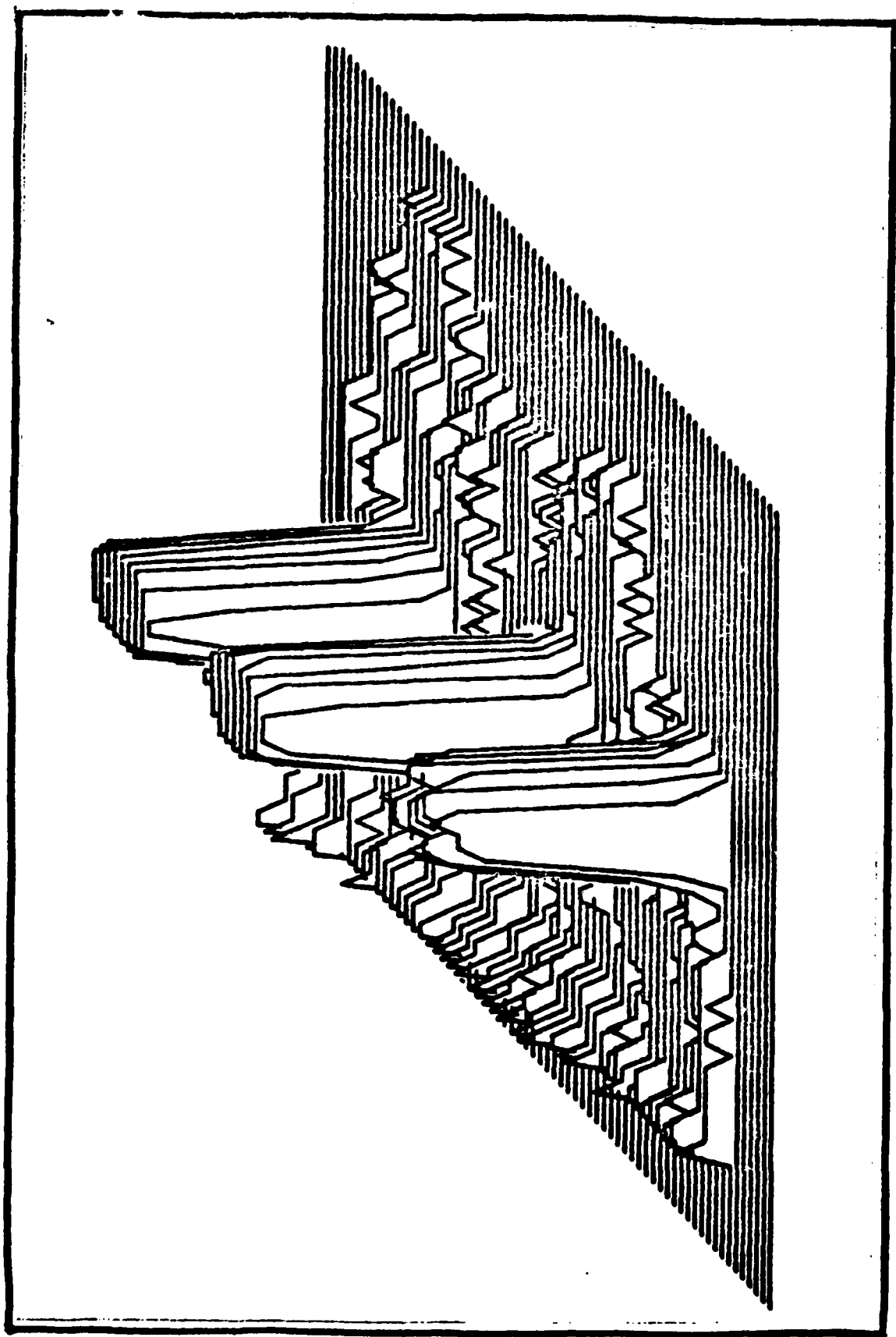


Fig. A.3. Template Density 3 with Cross-Entropy = 1.29606

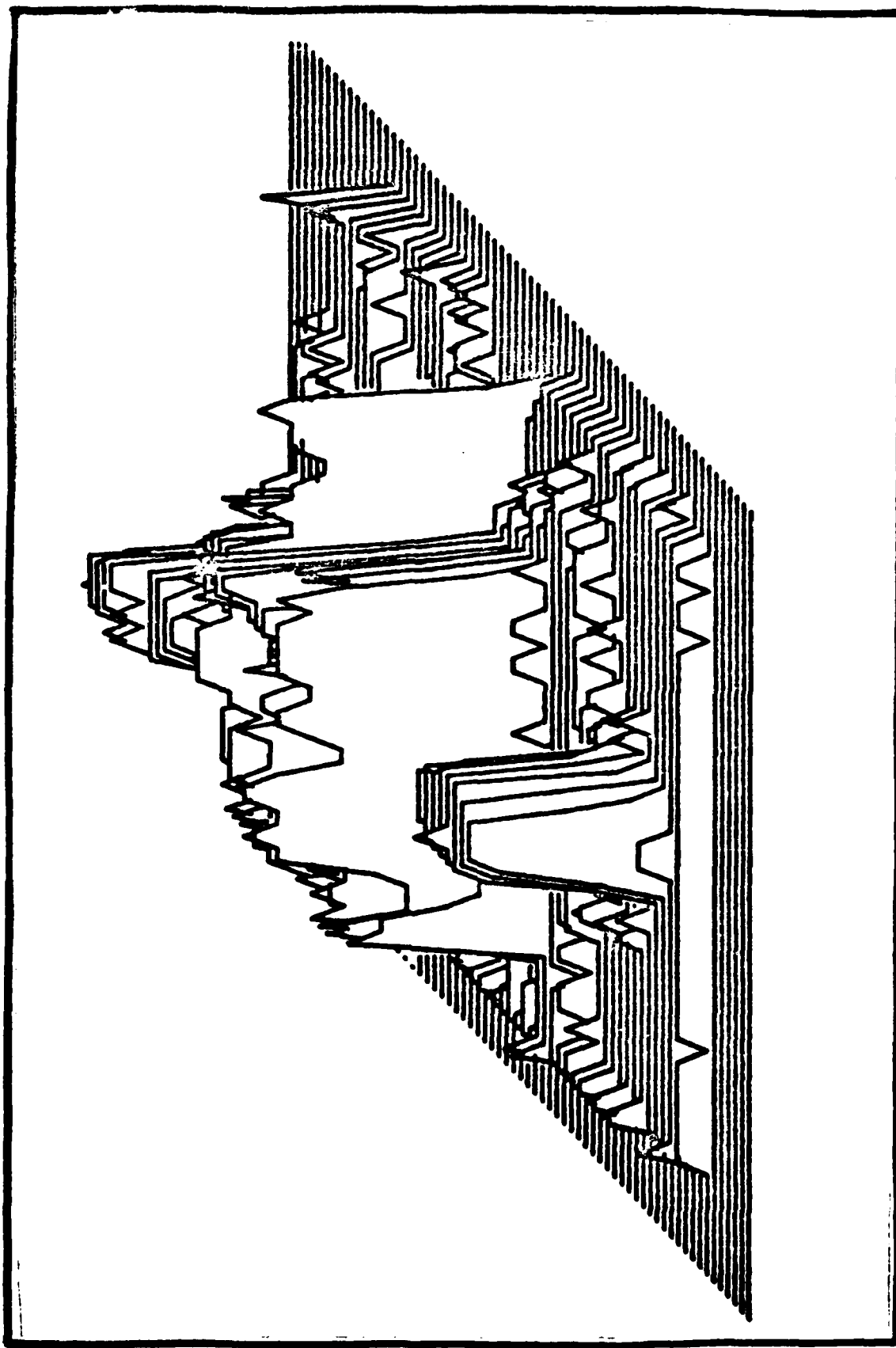


Fig. A.4. Template Density 4 with Cross-Entropy = 0.93407

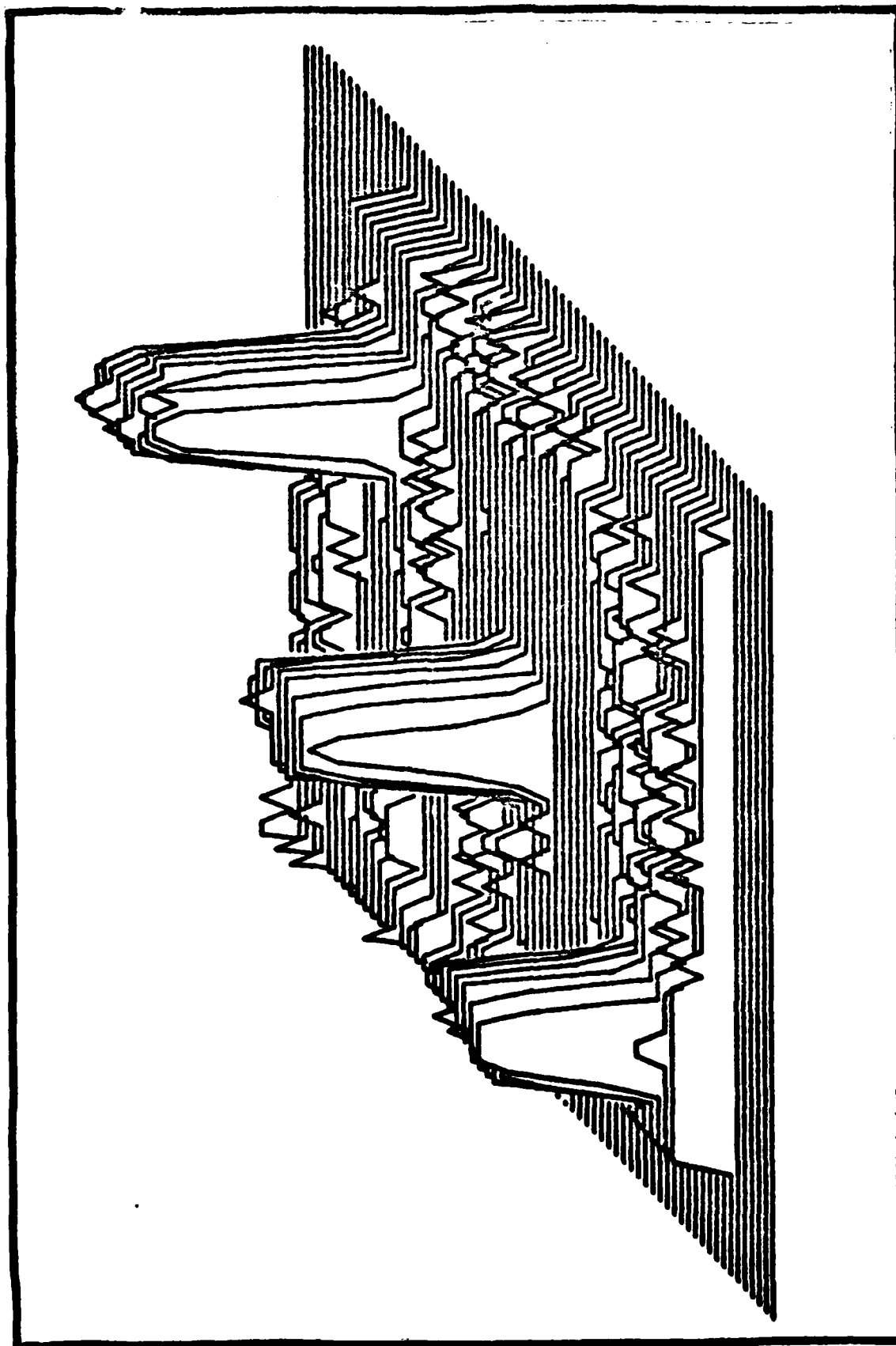


Fig. A.5. Template Density 5 with Cross-Entropy = 0.79444

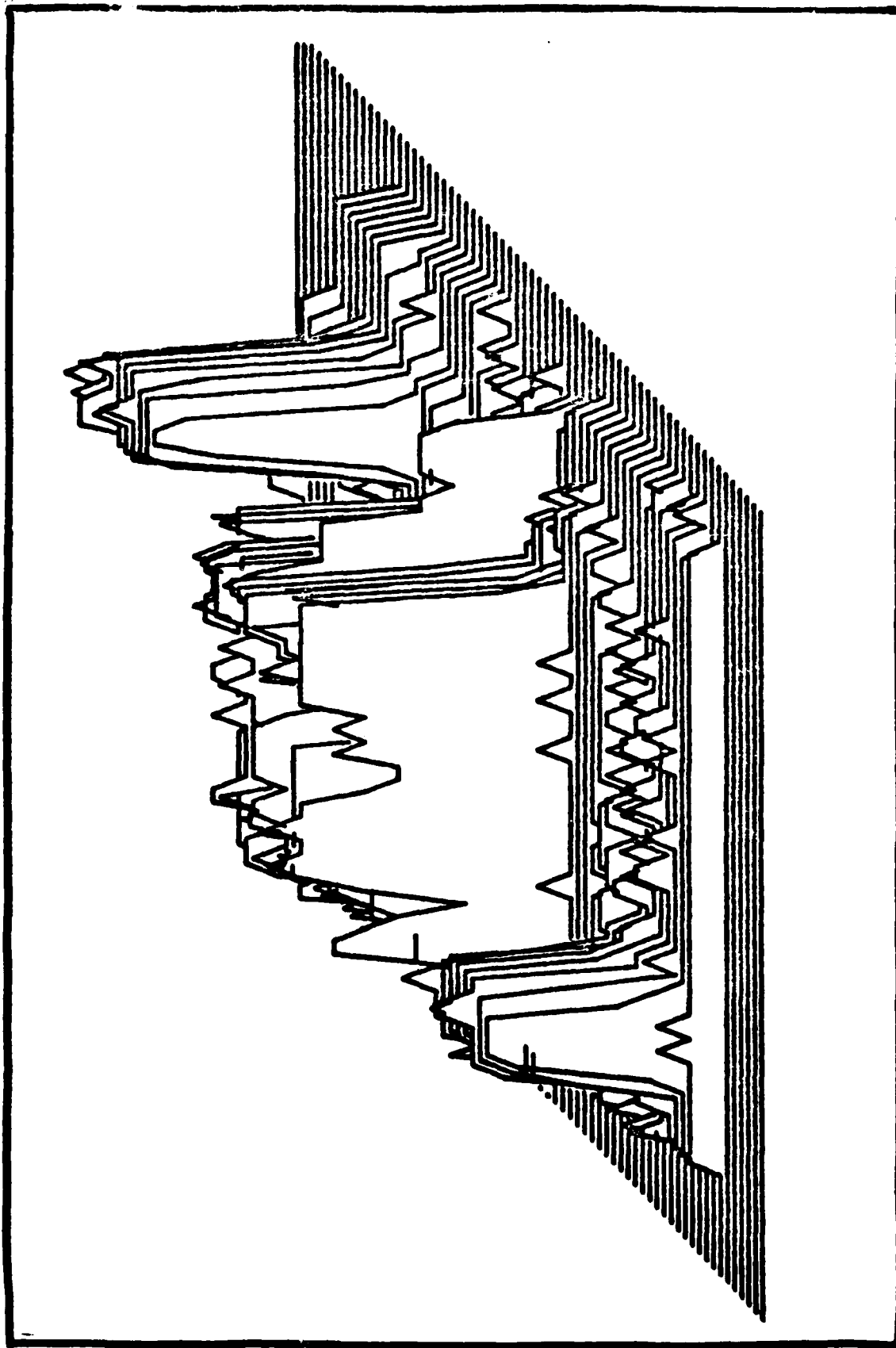


Fig. A.6. Template Density 6 with Cross-Entropy = 0.90418

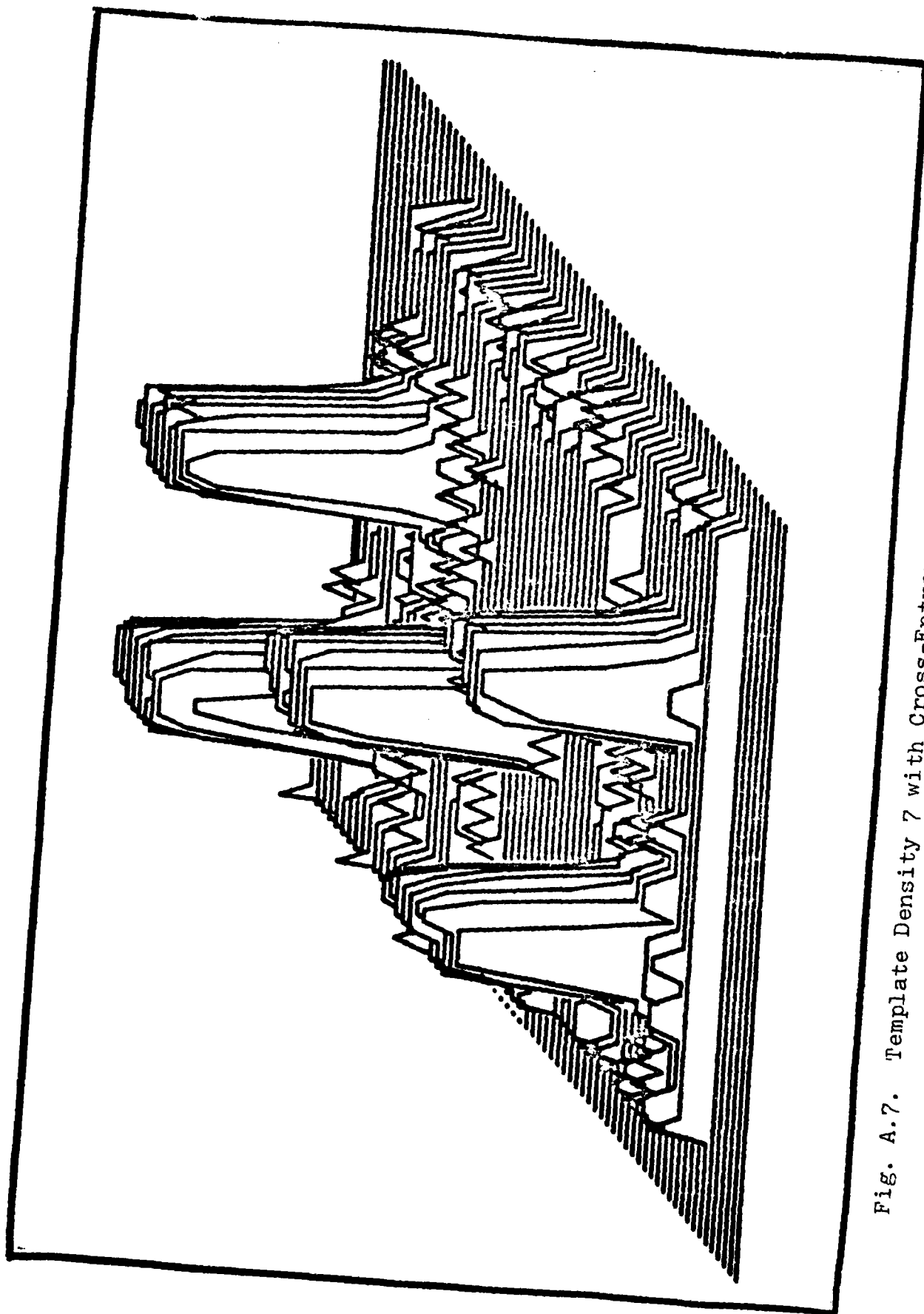


Fig. A.7. Template Density 7 with Cross-Entropy = 0.87136



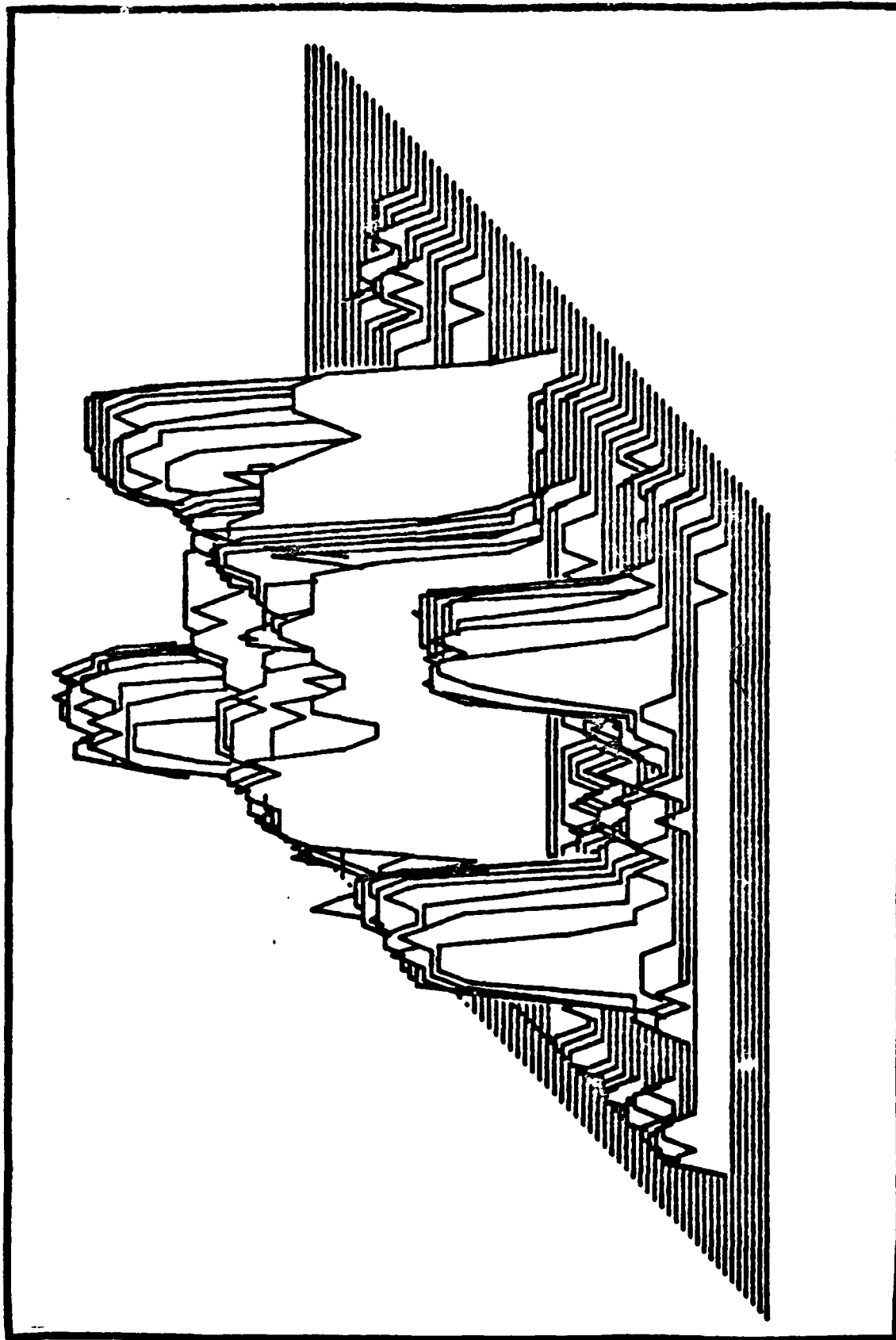


Fig. A.8. Template Density 8 with Cross-Entropy = 0.96869

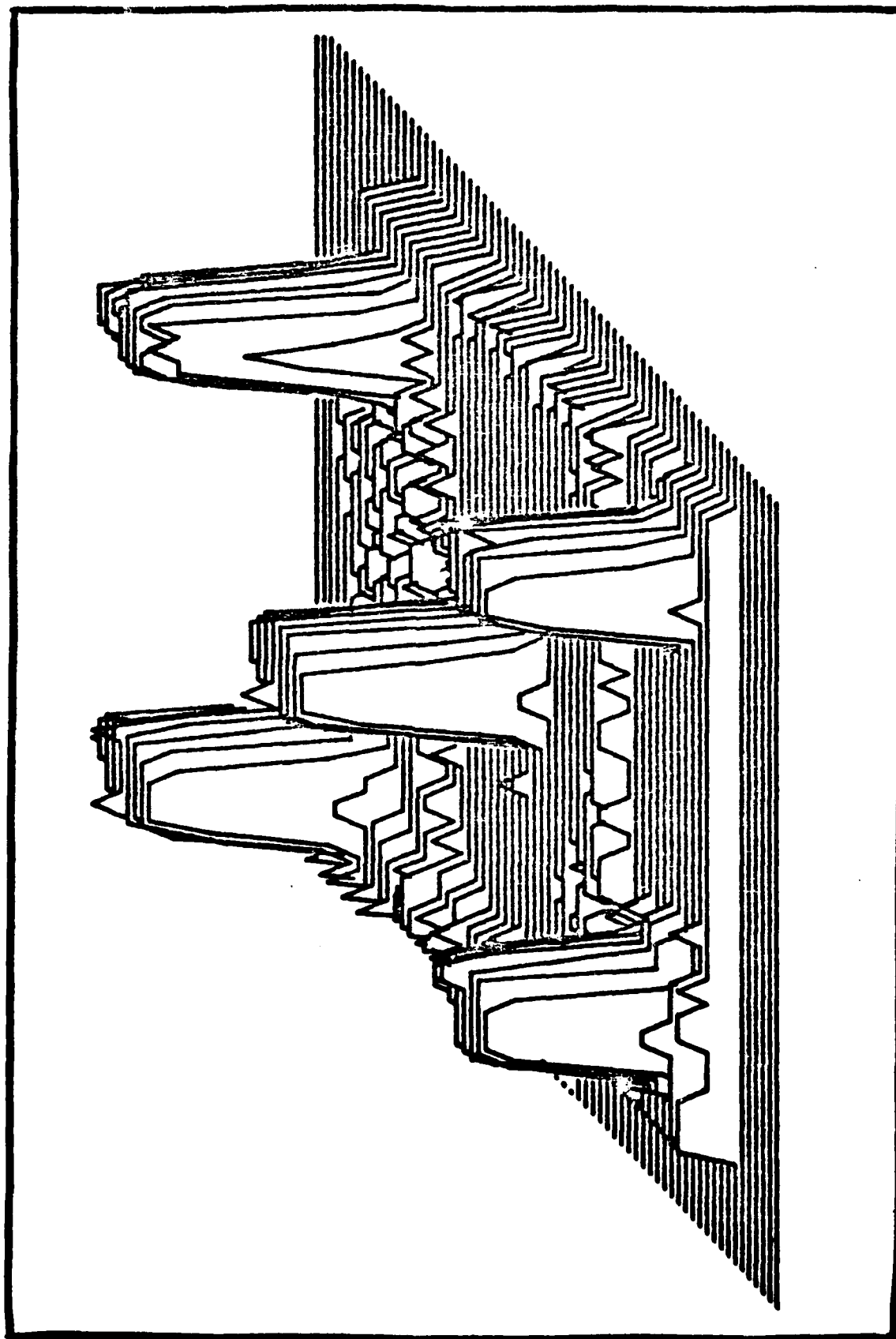


Fig. A.9. Template Density 9 with Cross-Entropy = 0.86579

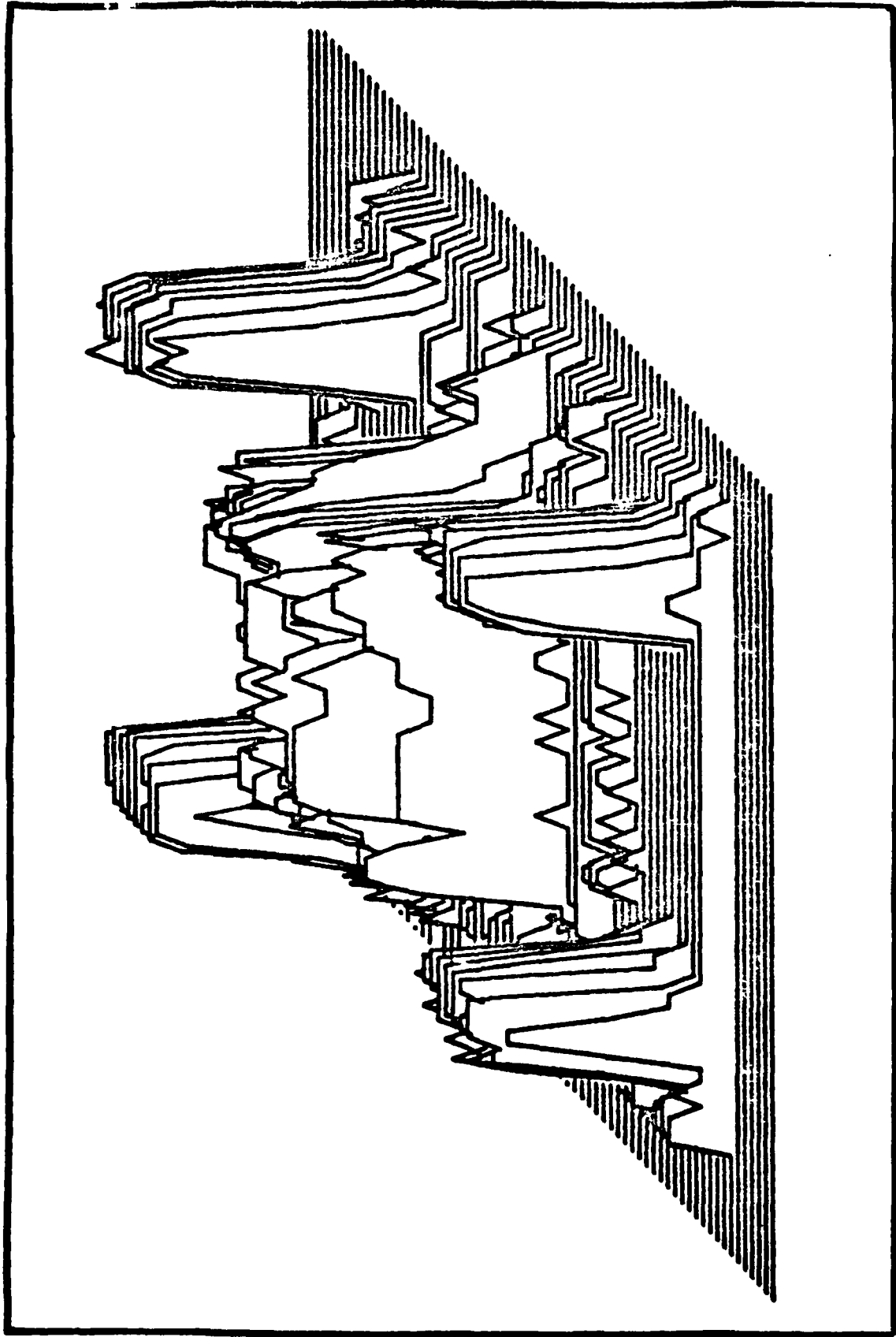


Fig. A.10. Template Density 10 with Cross-Entropy = 0.91397

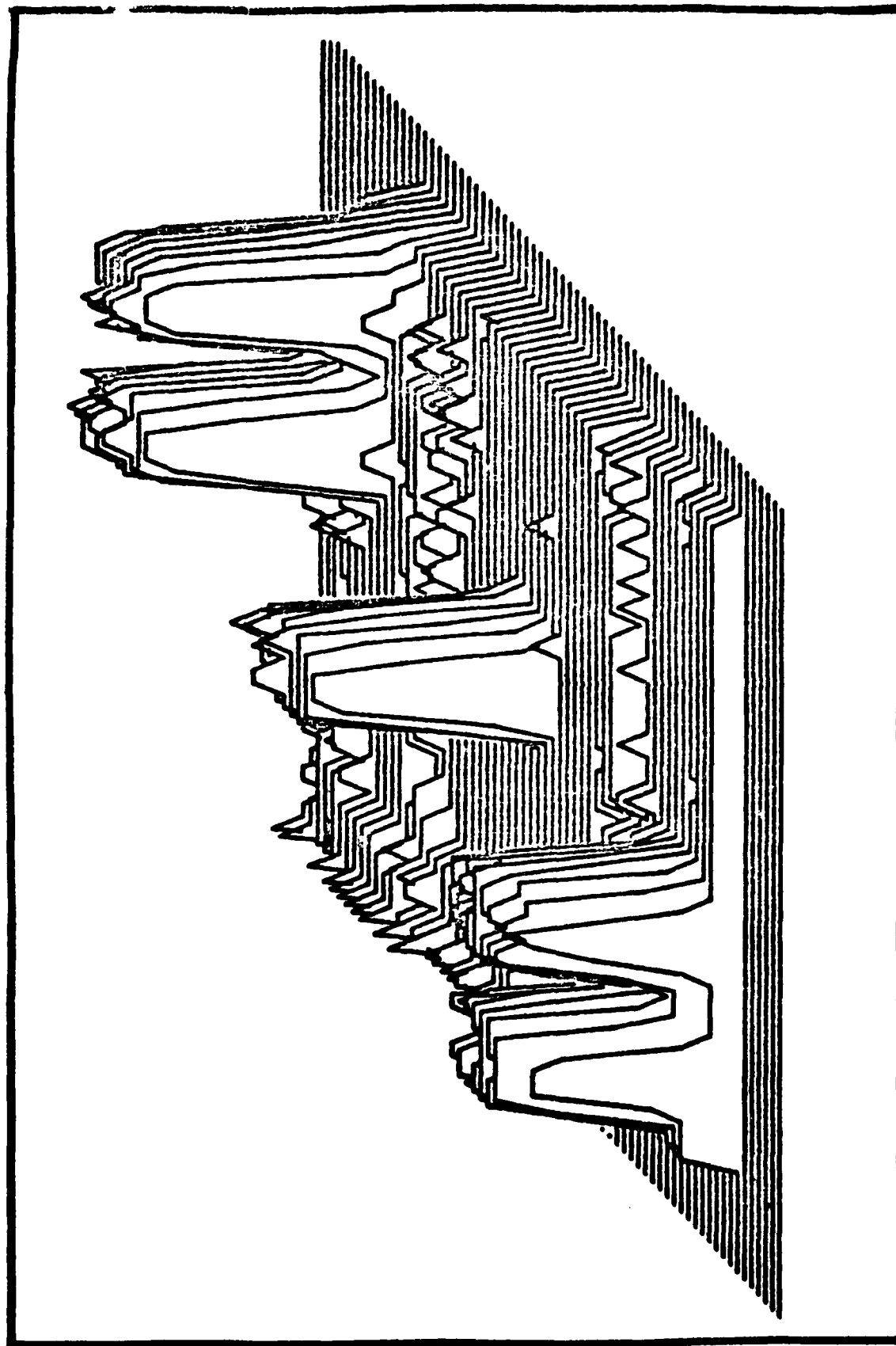


Fig. A.11. Template Density 11 with Cross-Entropy = 0.85447

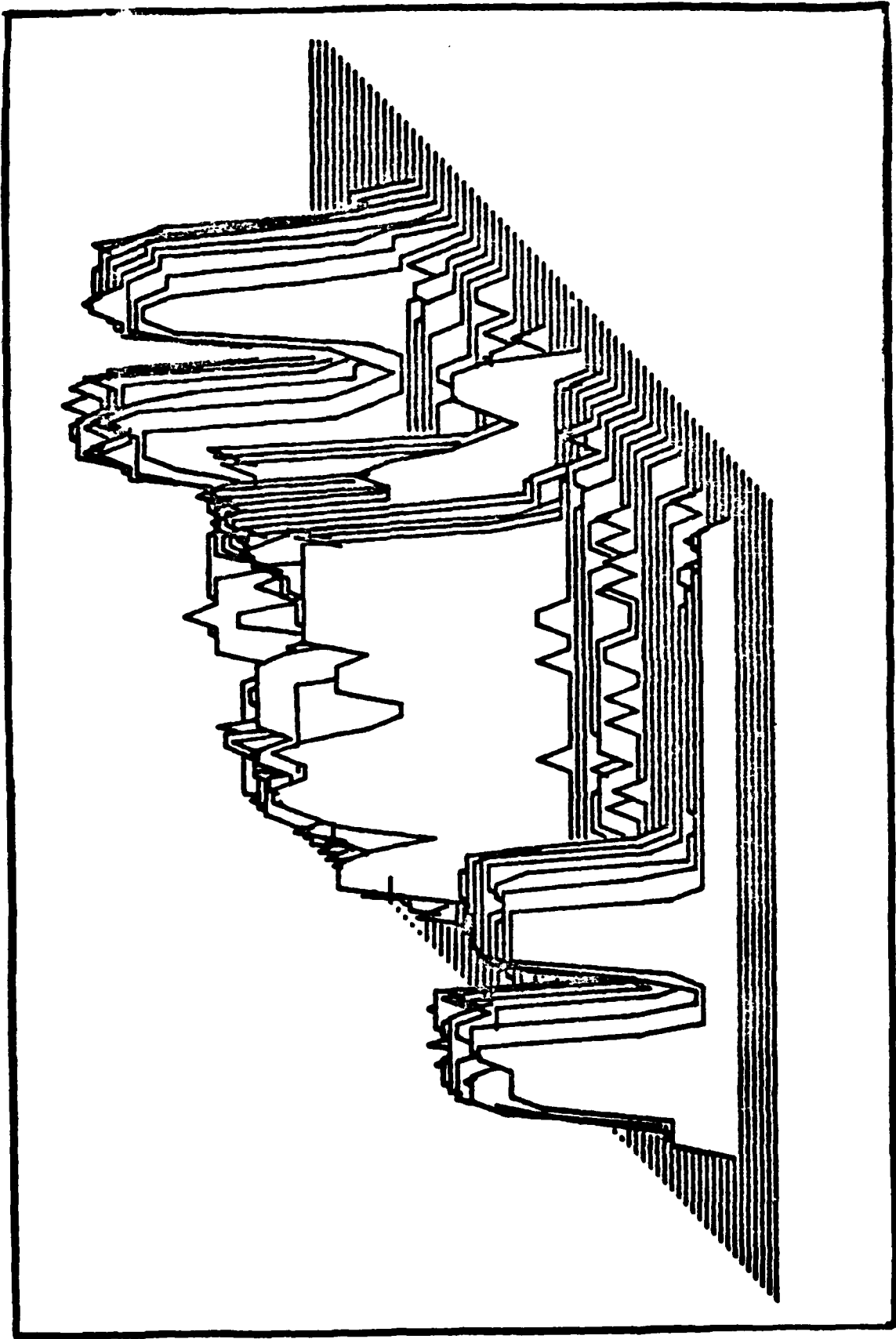


Fig. A.12. Template Density 12 with Cross-Entropy = 0.92389

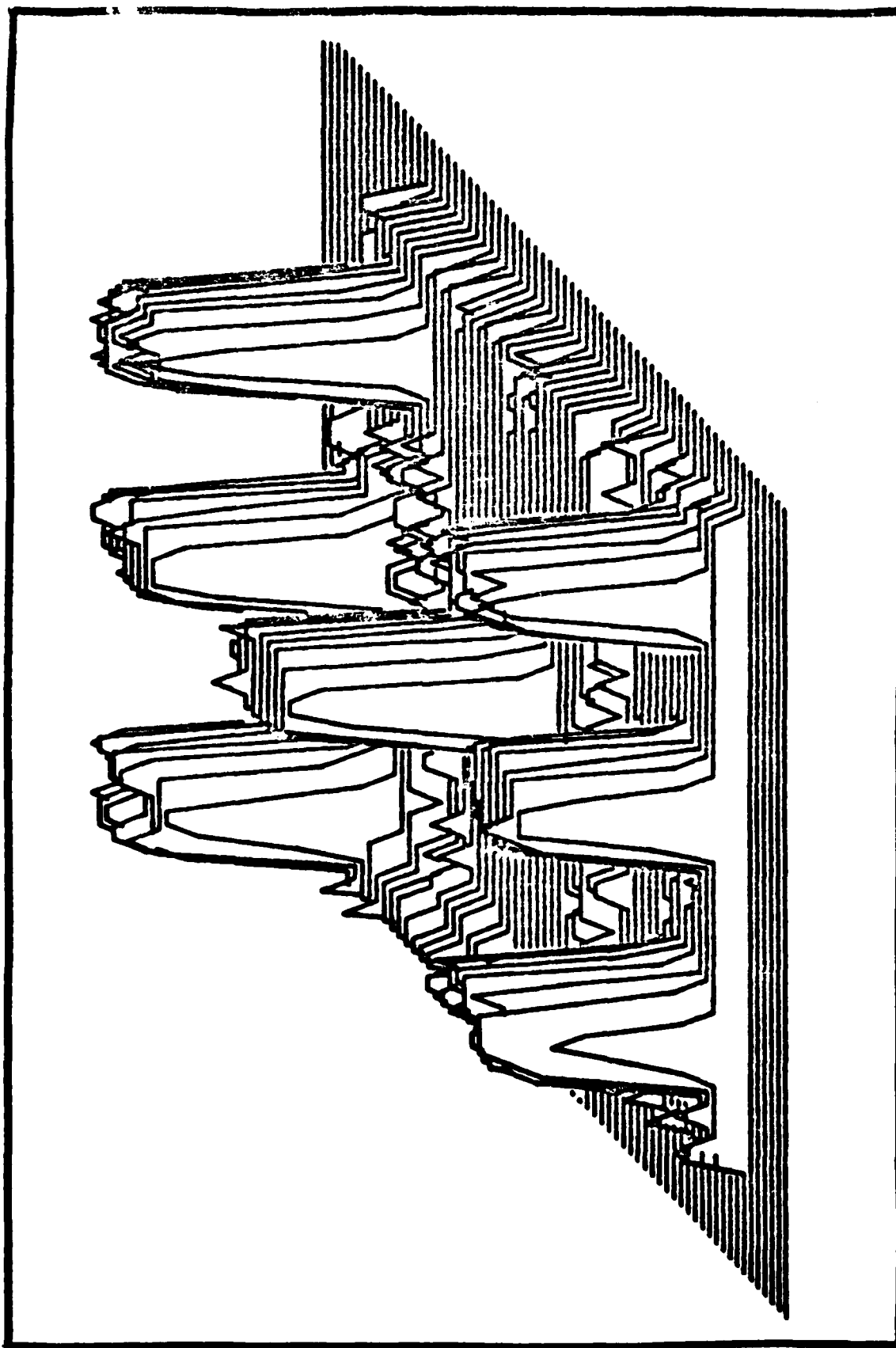


Fig. A.13. Template Density 13 with Cross-Entropy = 0.88835

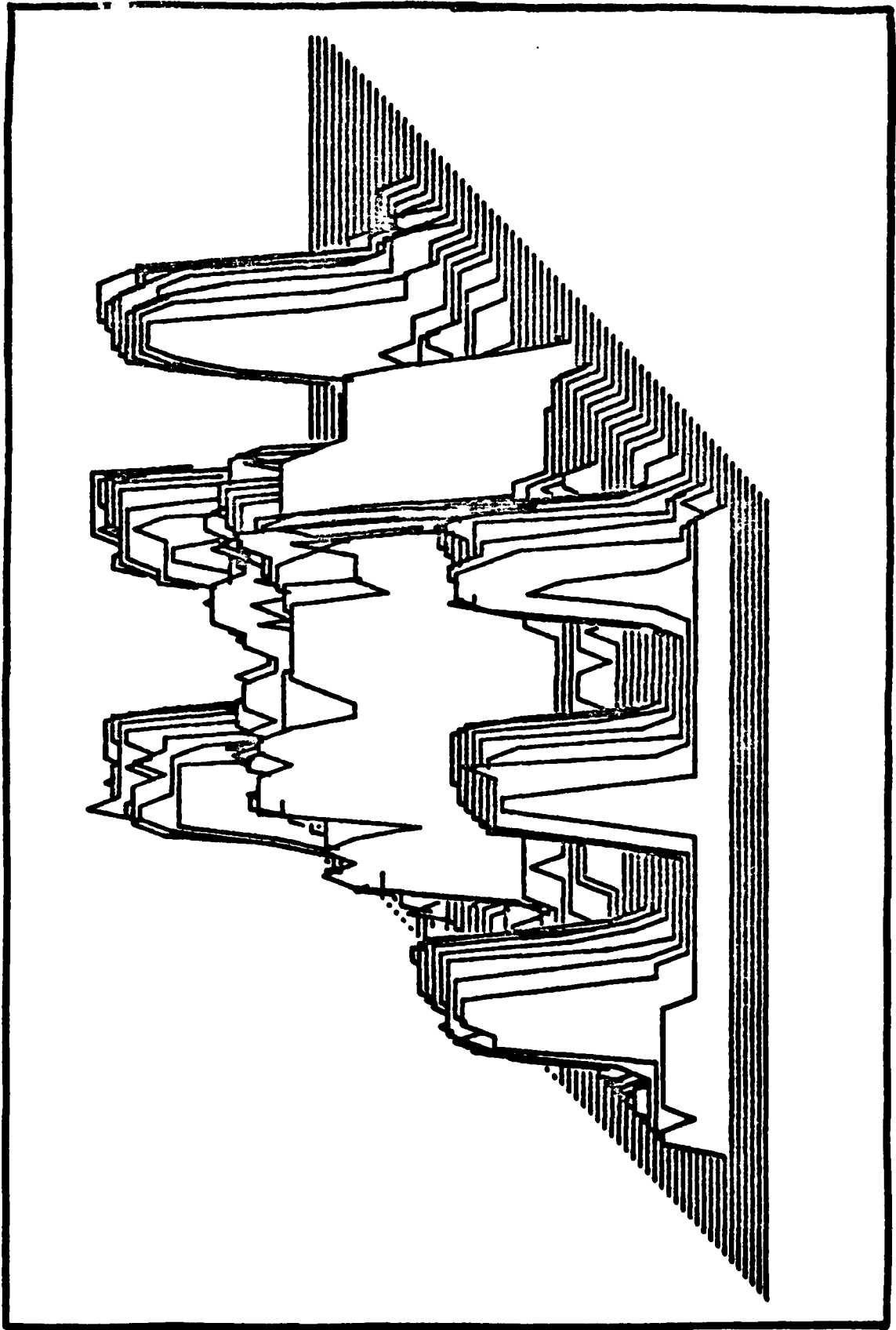


Fig. A.14. Template Density 14 with Cross-Entropy = 0.91680

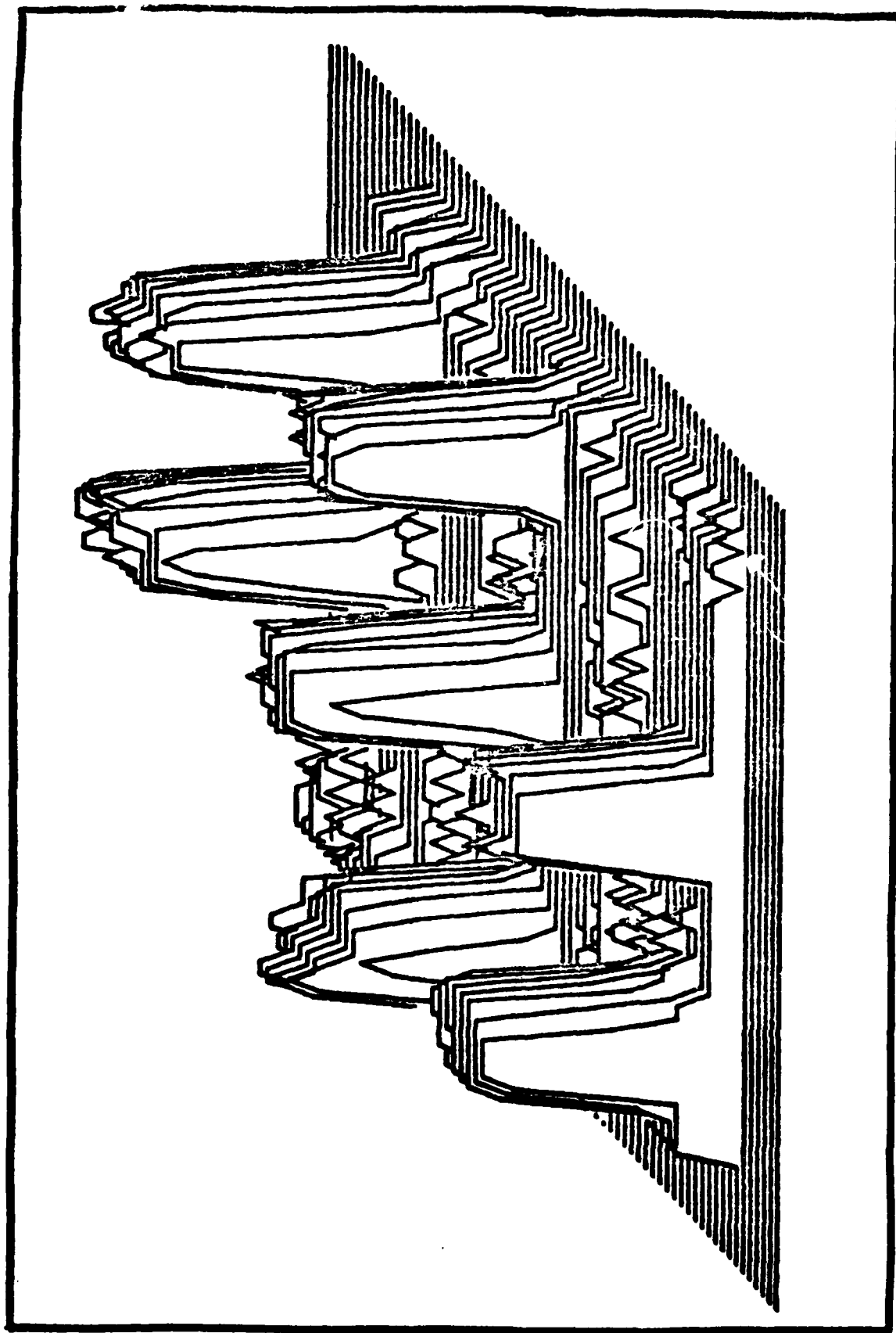


Fig. A.15. Template Density 15 with Cross-Entropy = 0.88692



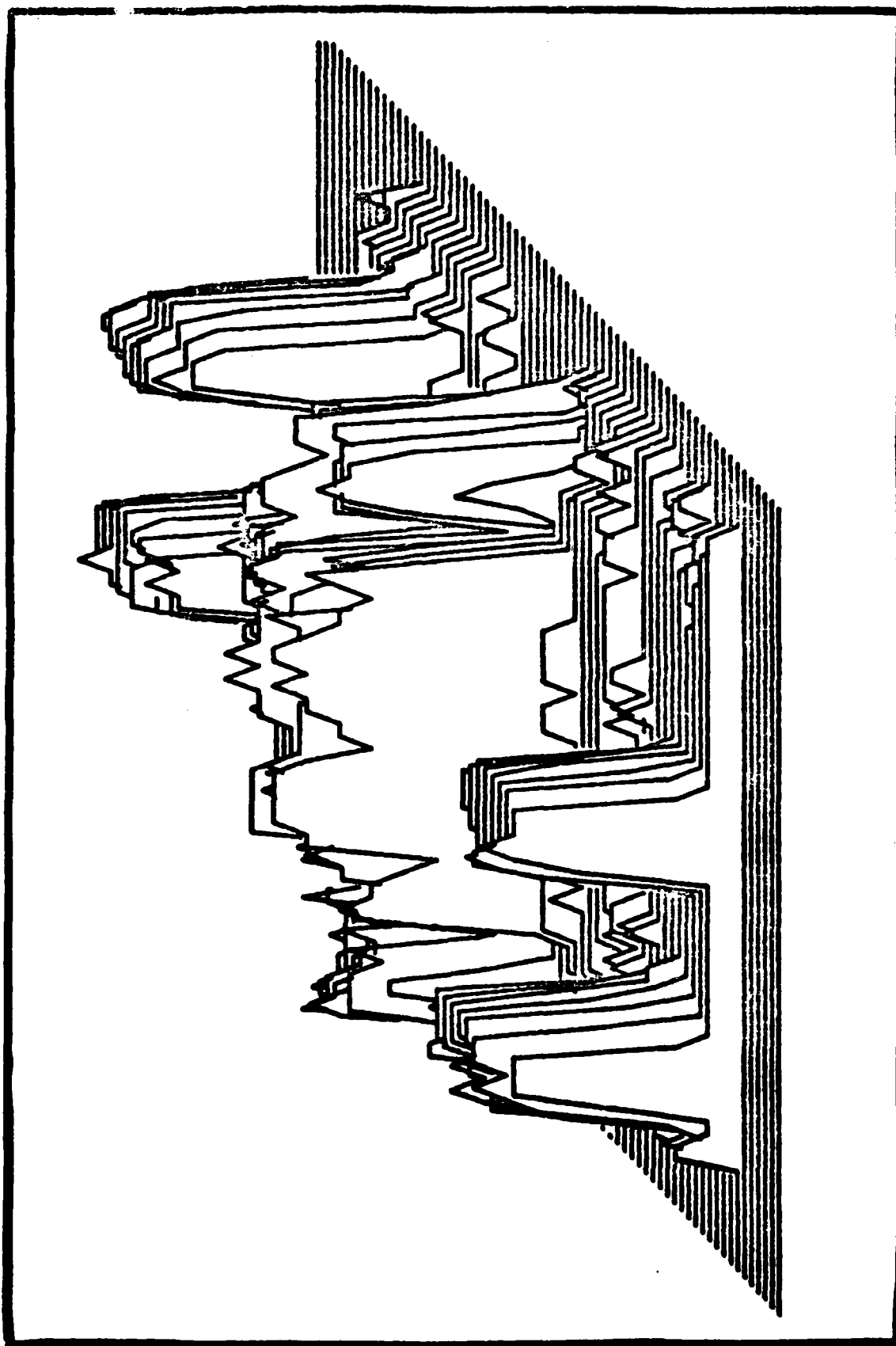


Fig. A.16. Template Density 16 with Cross-Entropy = 0.91118

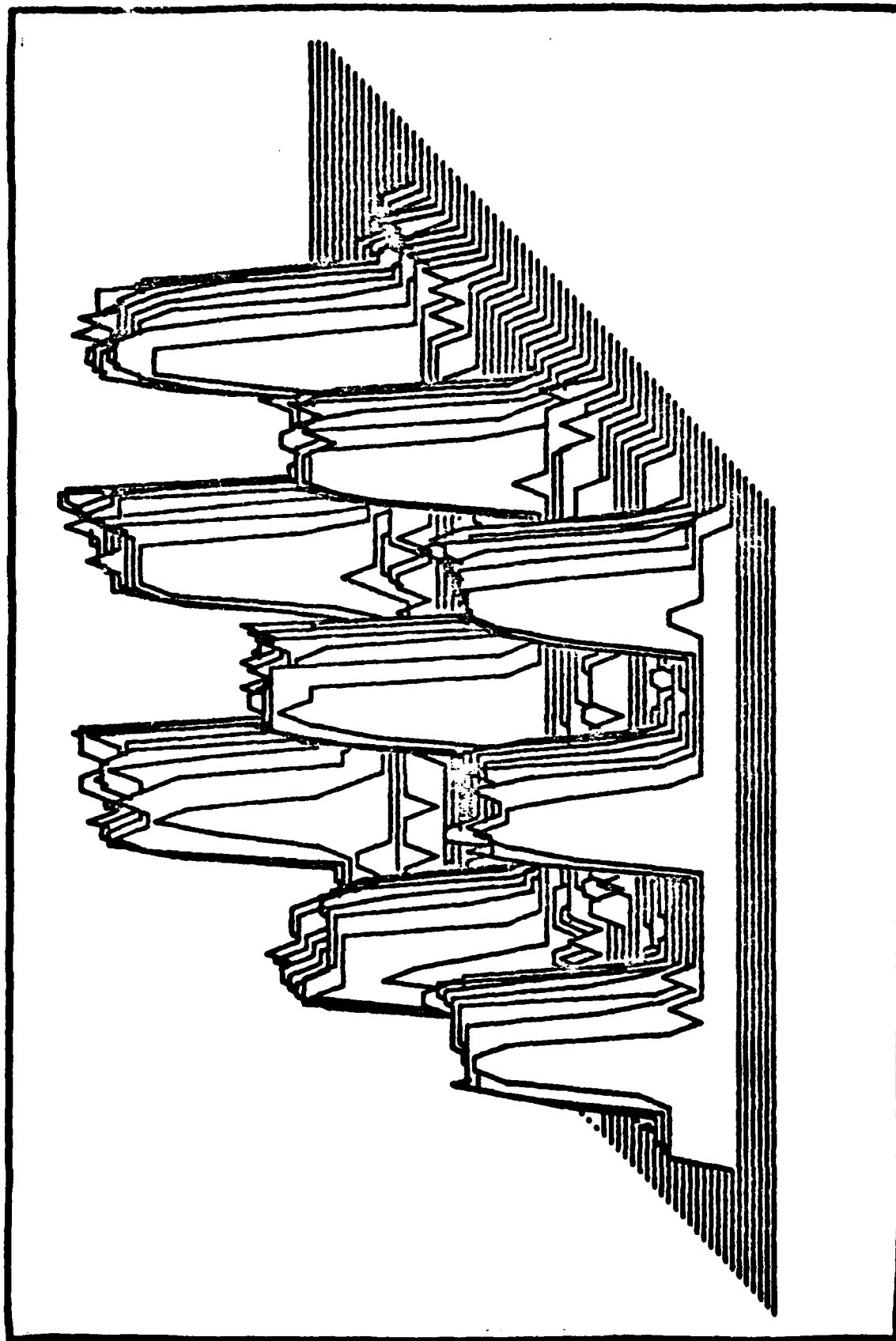


Fig. A.17. Template Density 17 with Cross-Entropy = 0.91693

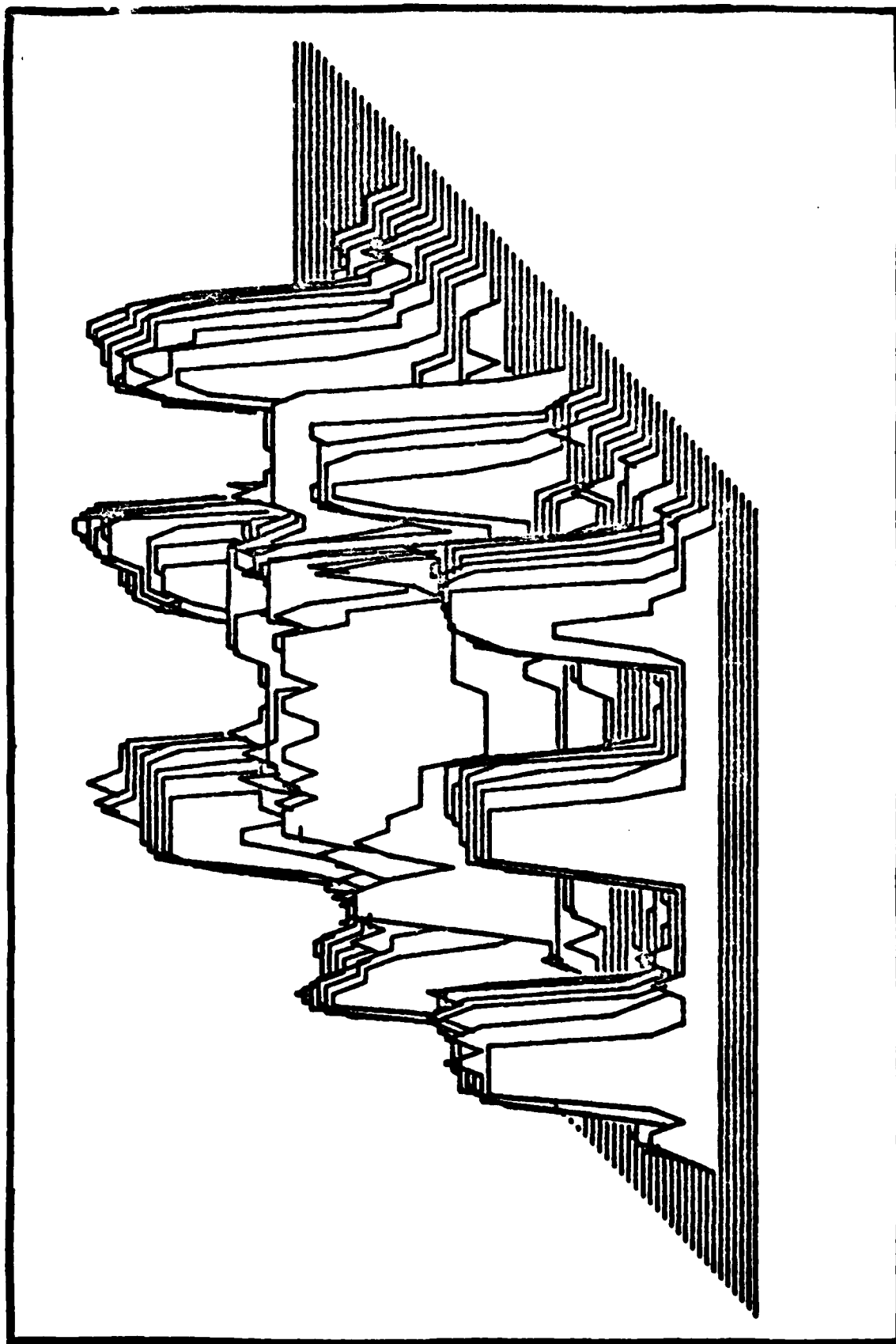


Fig. A.18. Template Density 18 with Cross-Entropy = 0.88837

## Appendix B. Minimum Cross-Entropy Templates

The eighteen minimum cross-entropy densities that correspond to the template scenes given in Appendix A are shown in this appendix. Given the lambda vector resulting from the solution of the constrained optimization problem defined by the template density moments, each of the minimum cross-entropy templates is completely defined by

$$t_k(x,y) = \exp \left[ -\lambda_0^{(k)} - \sum_{i=1}^{90} \lambda_i^{(k)} f_i(x,y) \right]$$

$k = 1 \dots 18 \quad (B.1)$

Note the very low cross-entropy values resulting from the camera noise.

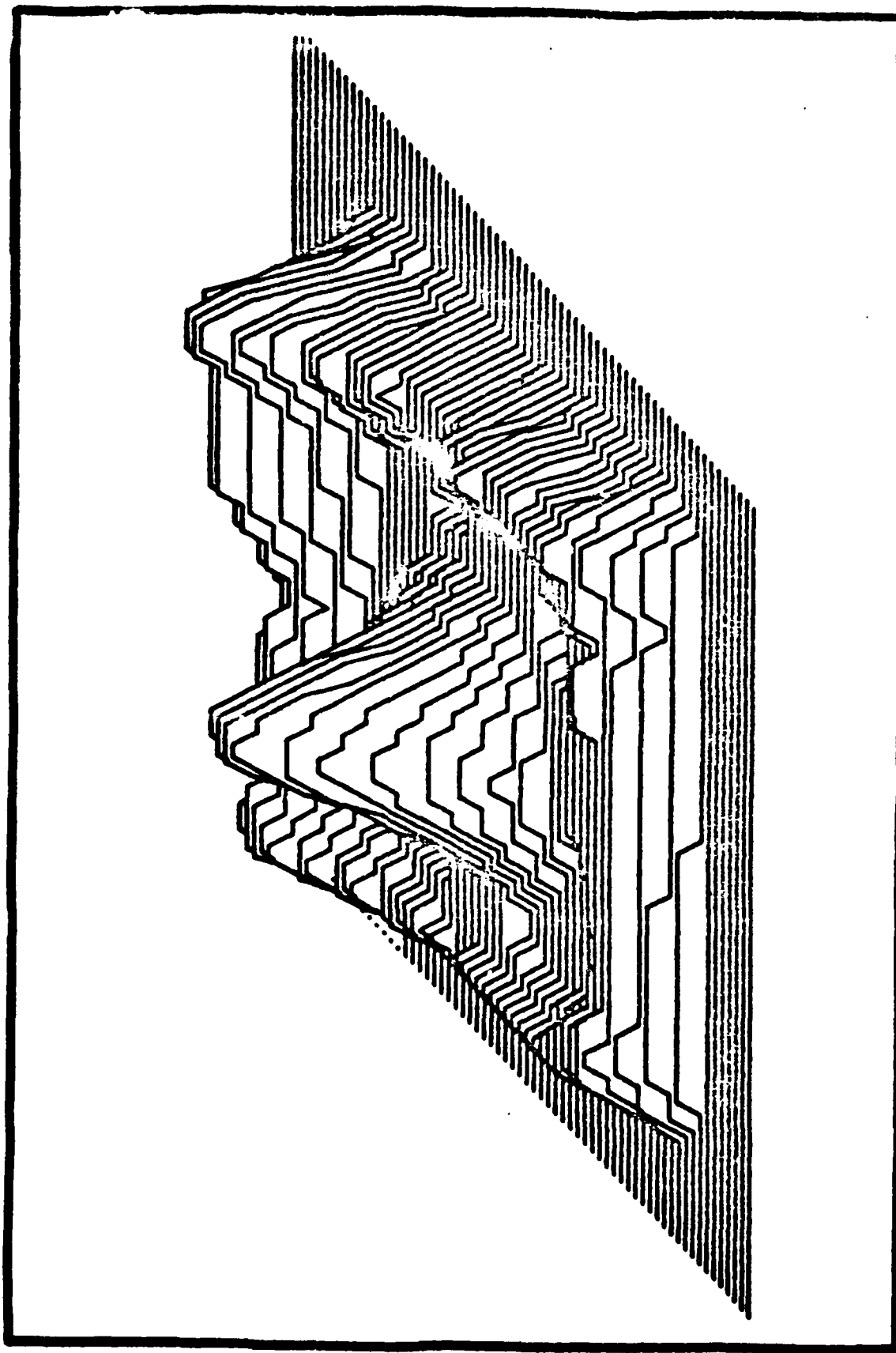


Fig. B.1. Minimum Cross-Entropy Template 1 with Cross-Entropy = 0.34217

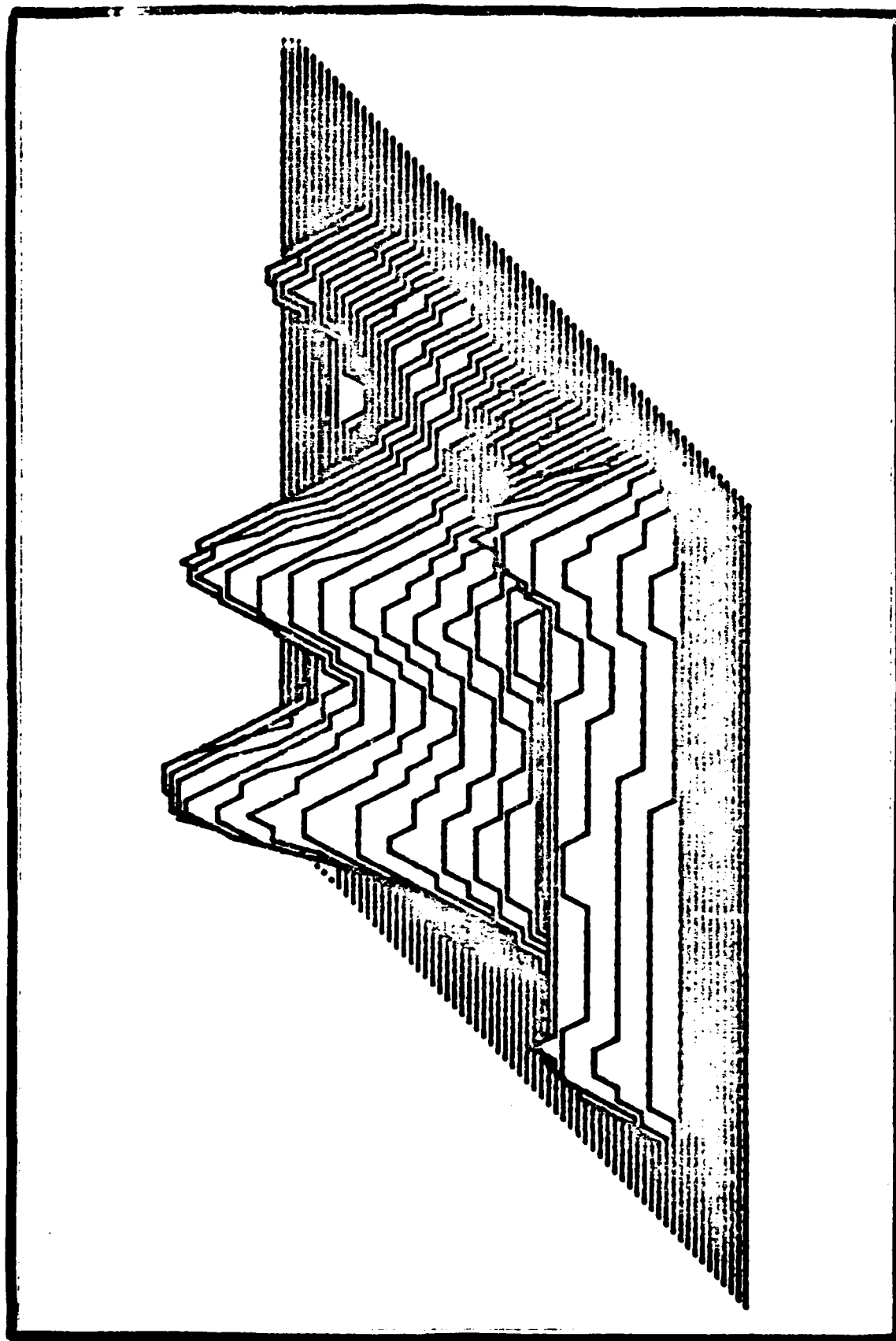


Fig. B.2. Minimum Cross-Entropy Template 2 with Cross-Entropy = 0.40756

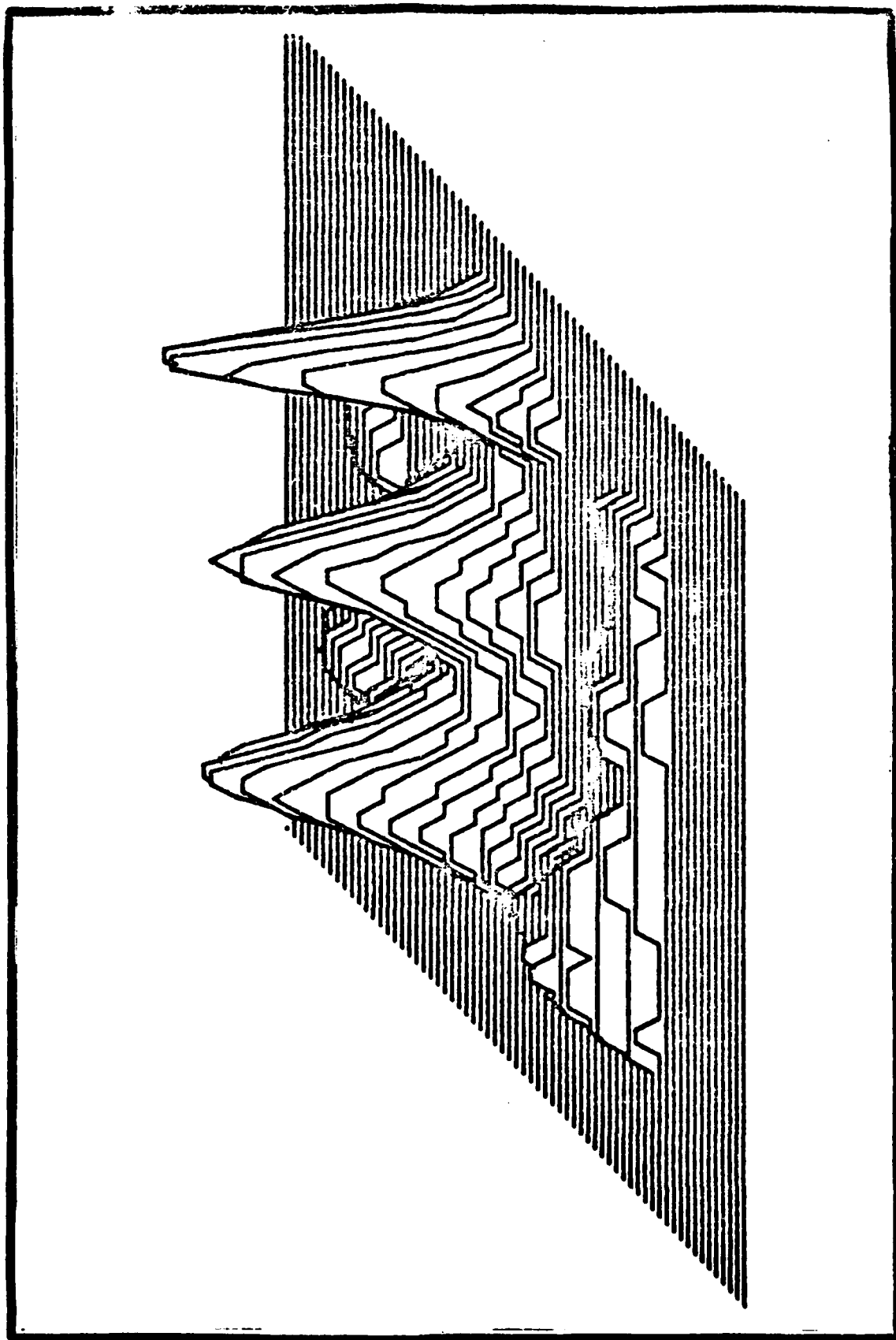


Fig. B.3. Minimum Cross-Entropy Template 3 with Cross-Entropy = 0.44235

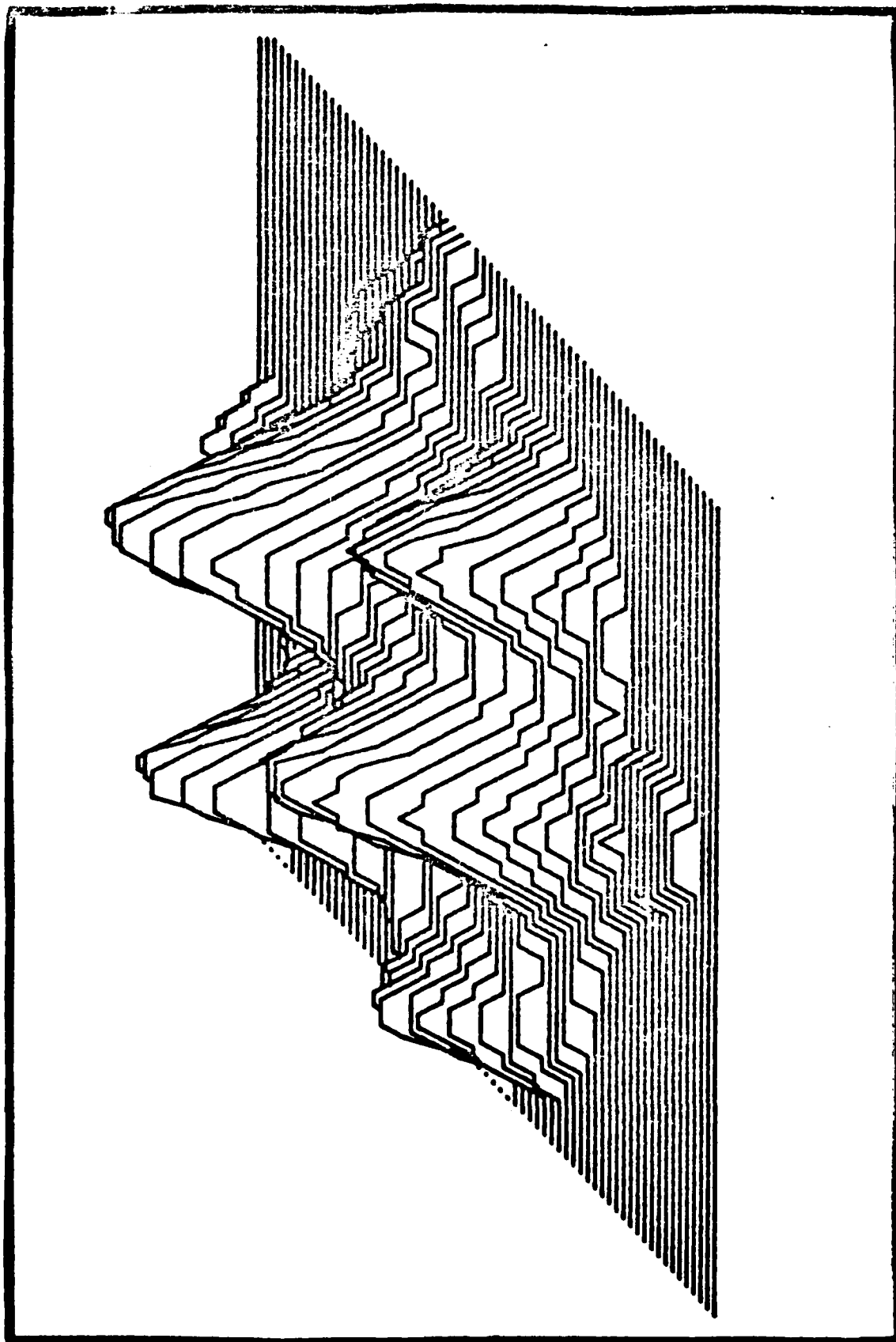


Fig. B.4. Minimum Cross-Entropy Template 4 with Cross-Entropy = 0.44962



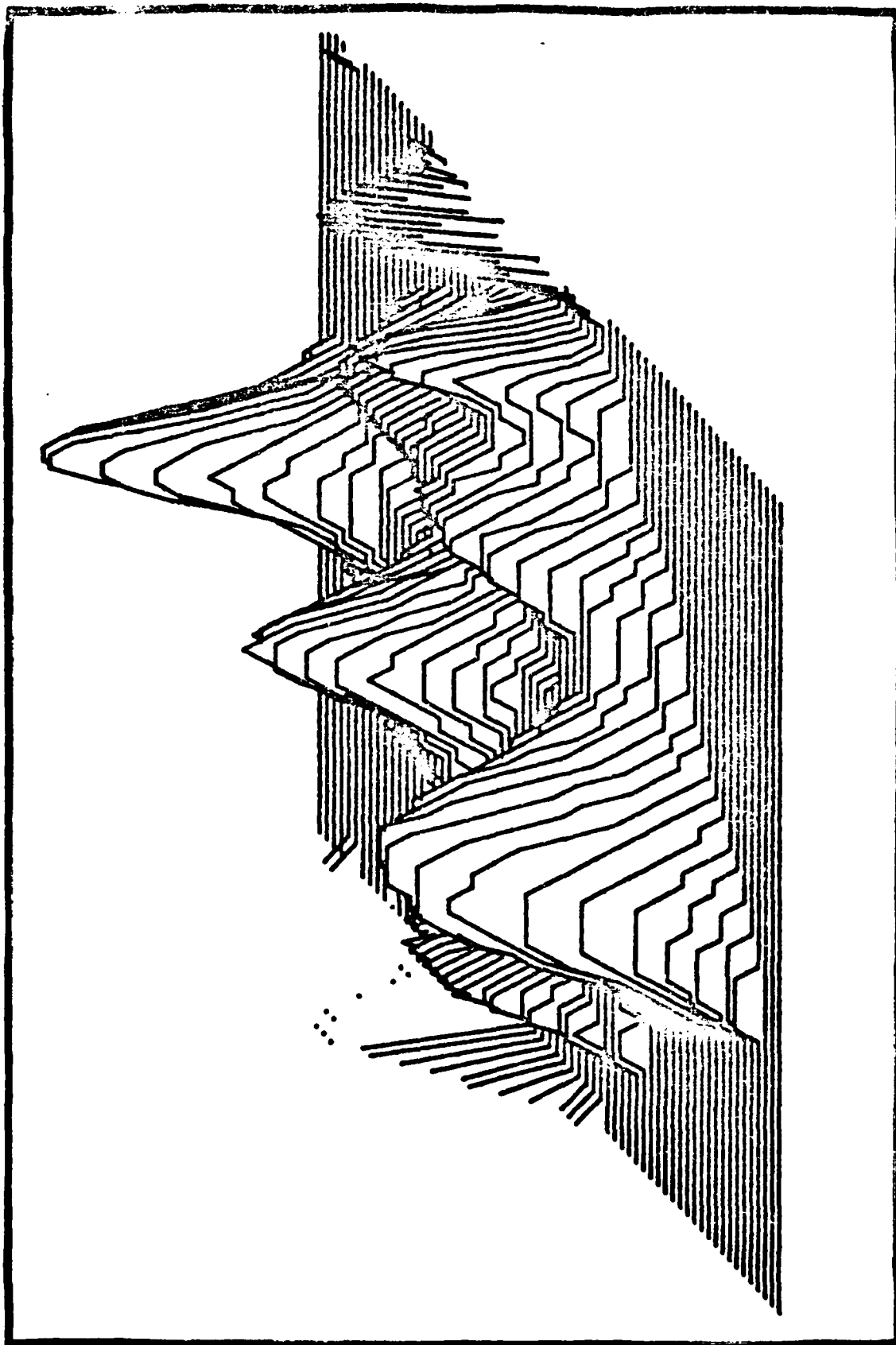


Fig. B.5. Minimum Cross-Entropy Template 5 with Cross-Entropy = 0.41877

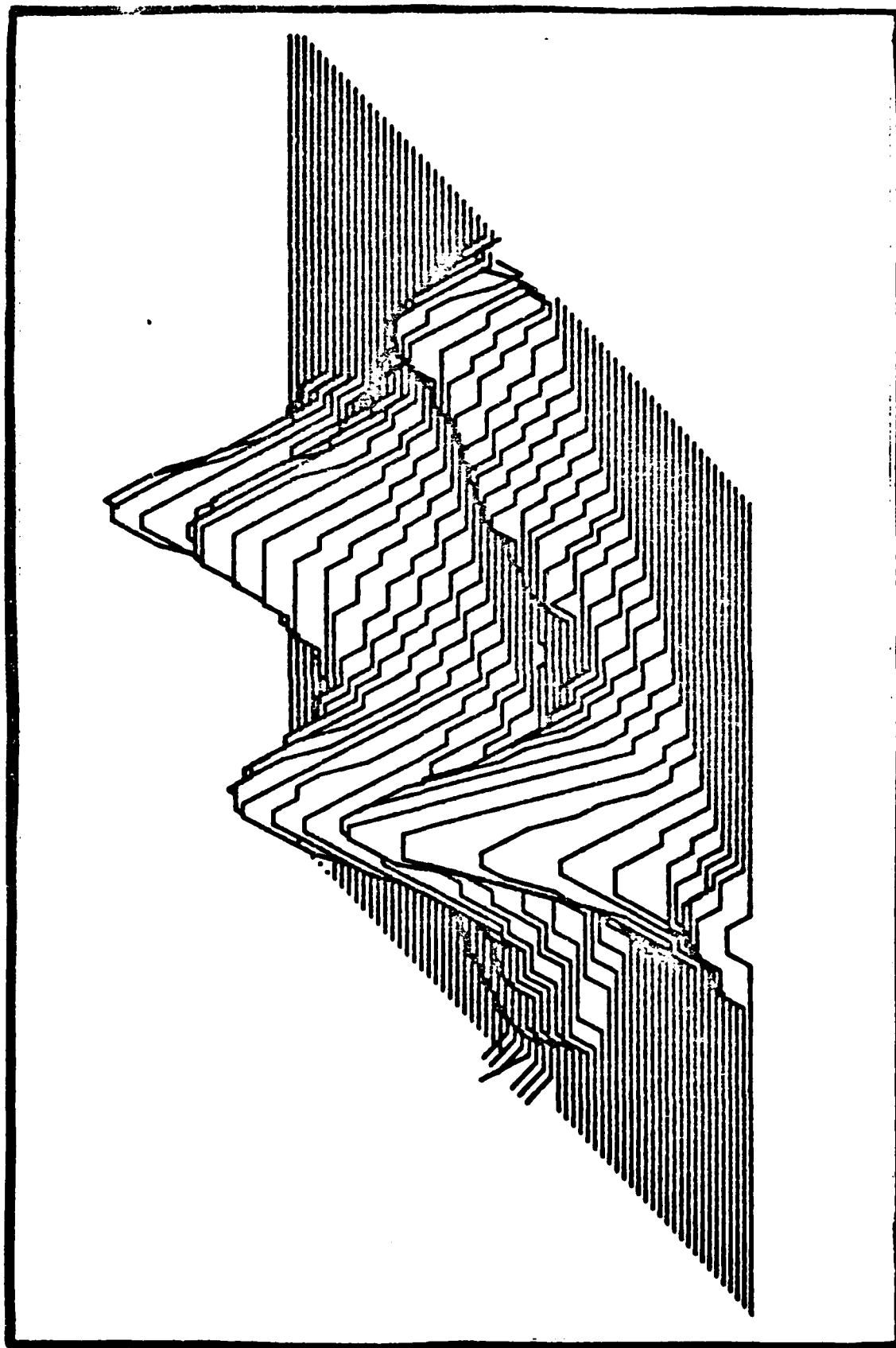


Fig. B.6. Minimum Cross-Entropy Template 6 with Cross-Entropy = 0.44274

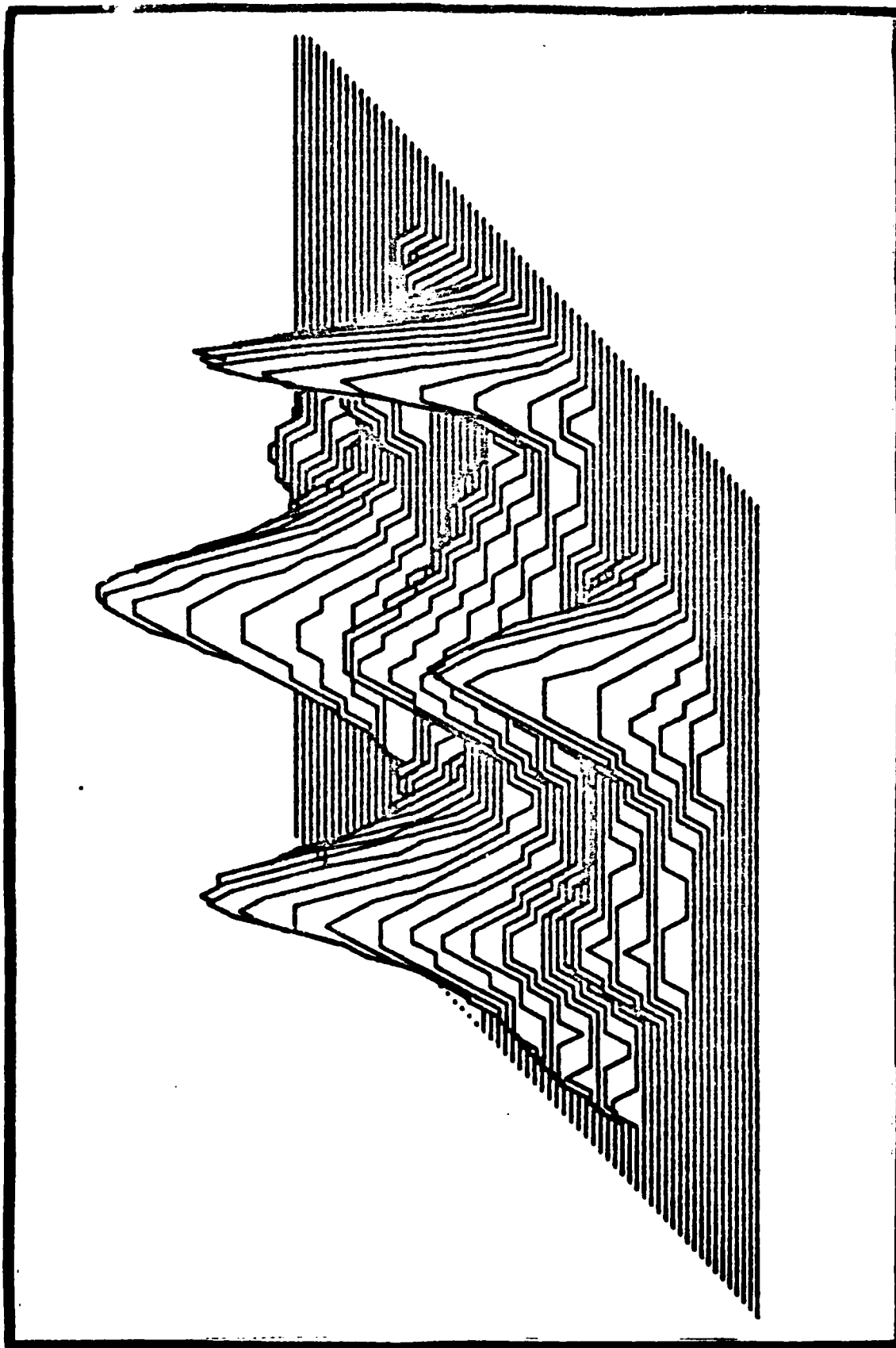


Fig. B.7. Minimum Cross-Entropy Template 7 with Cross-Entropy = 0.41746

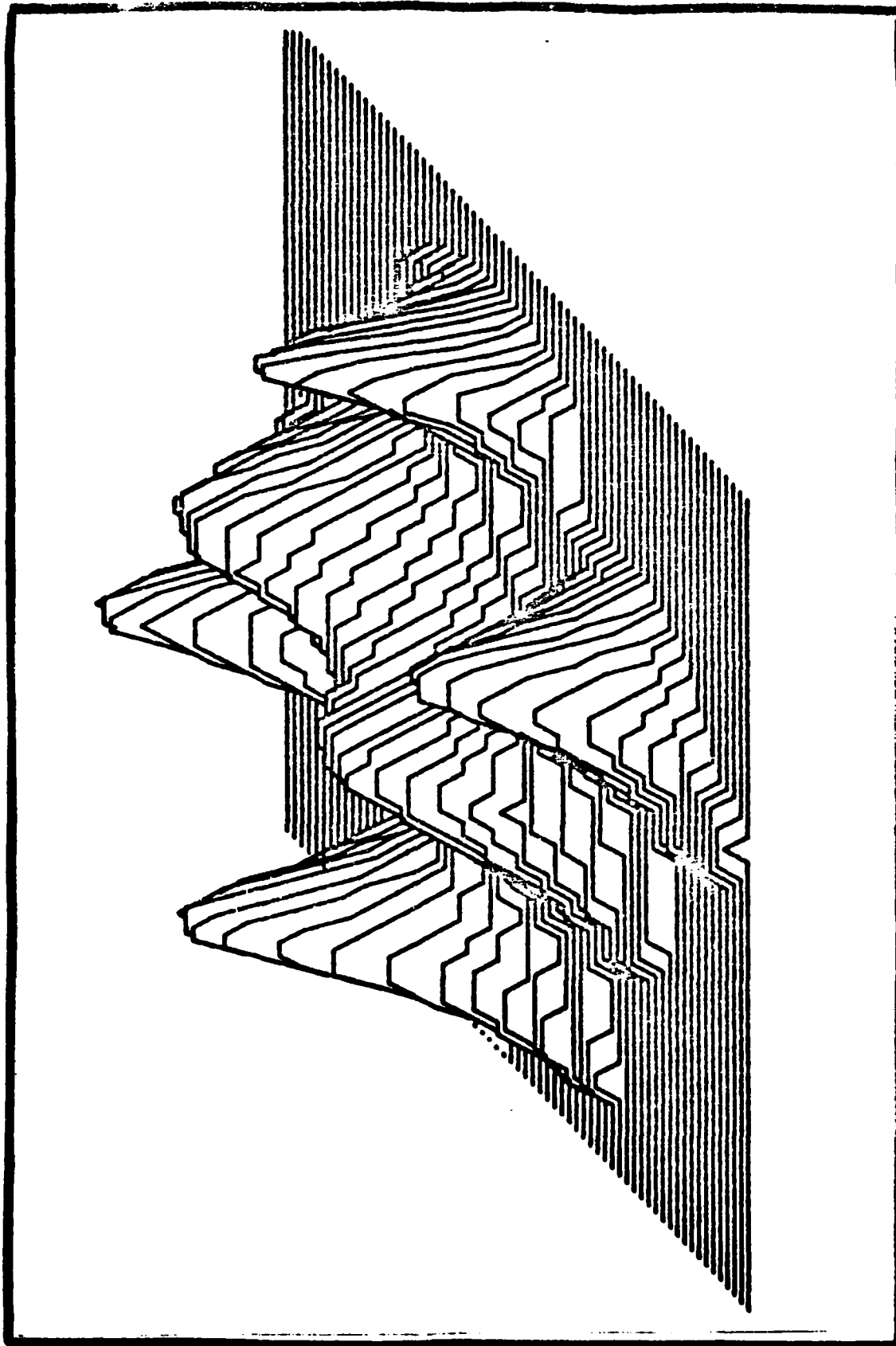


Fig. B.8. Minimum Cross-Entropy Template 8 with Cross-Entropy = 0.46145

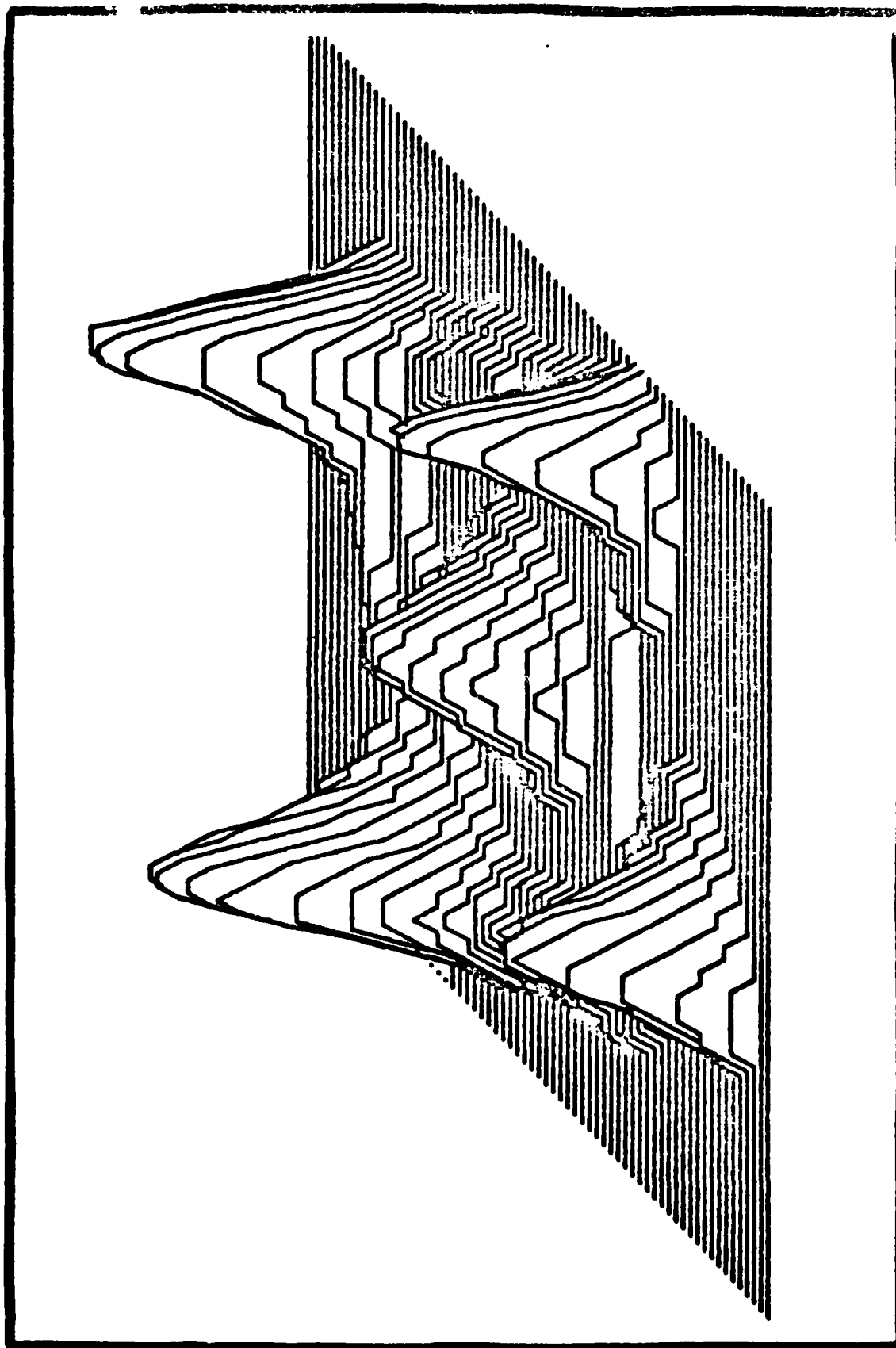


Fig. B.9. Minimum Cross-Entropy Template 9 with Cross-Entropy = 0.42566

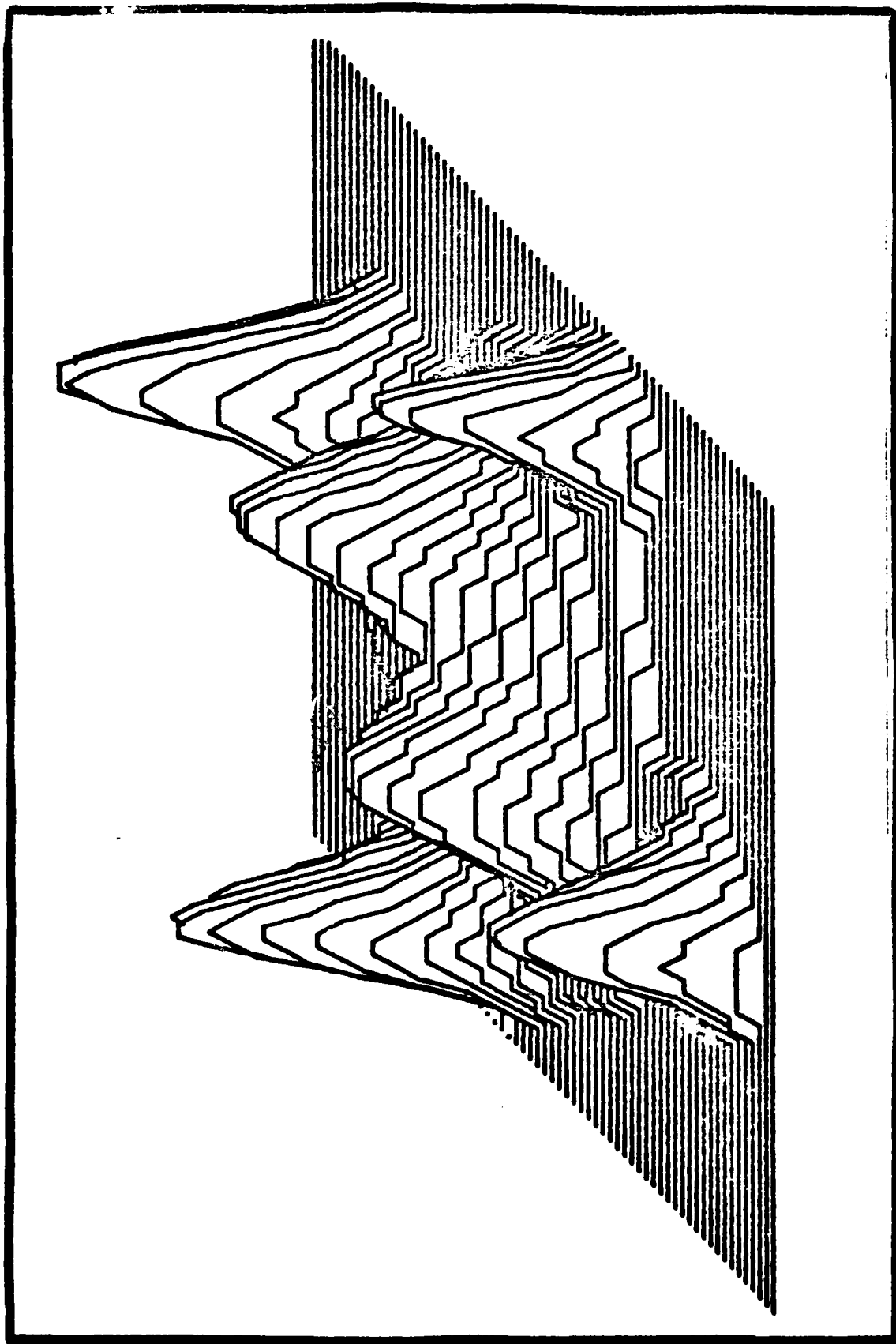


Fig. B.10. Minimum Cross-Entropy Template 10 with Cross-Entropy = 0.46862

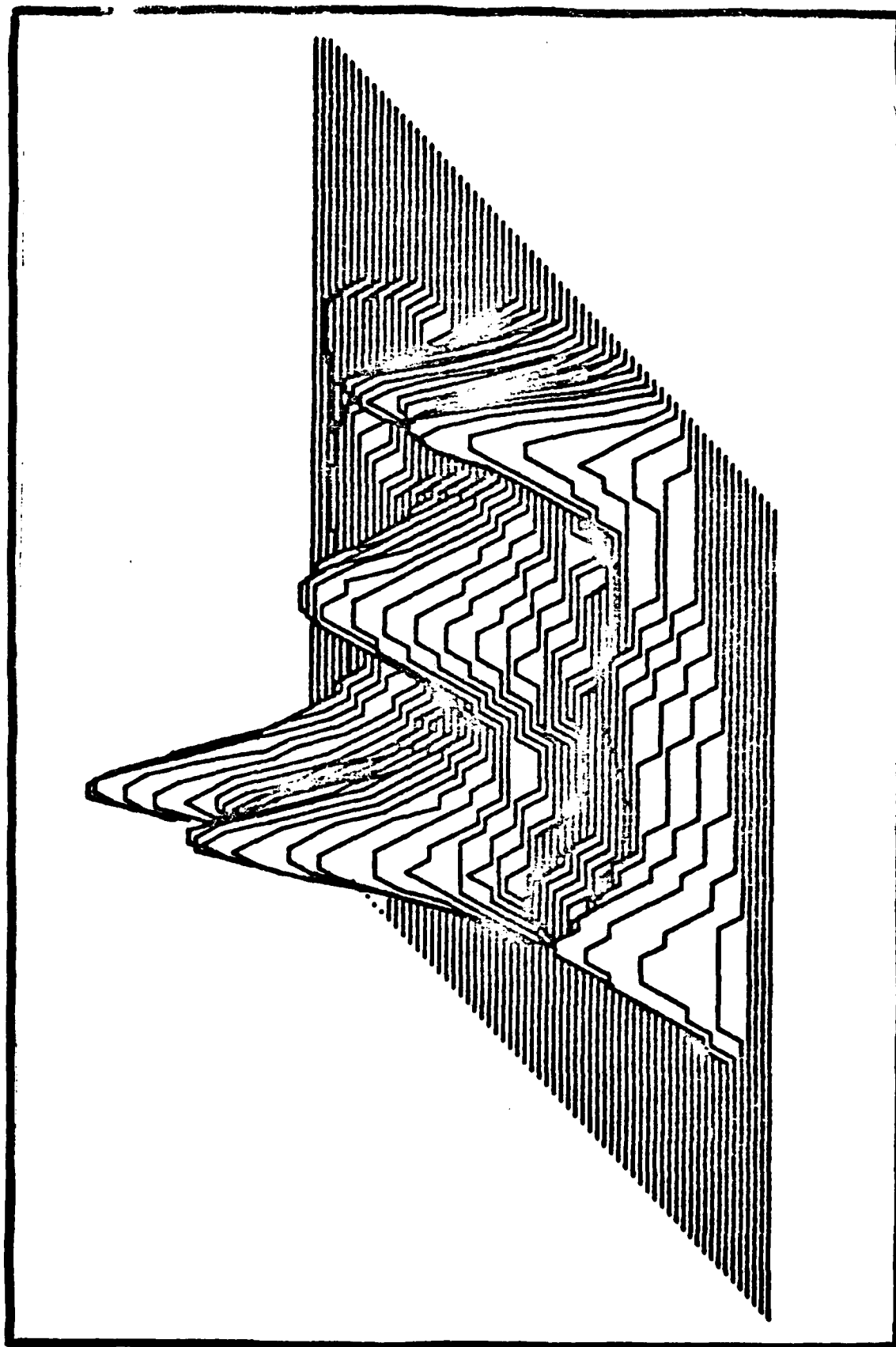


Fig. B.11. Minimum Cross-Entropy Template 11 with Cross-Entropy = 0.43961

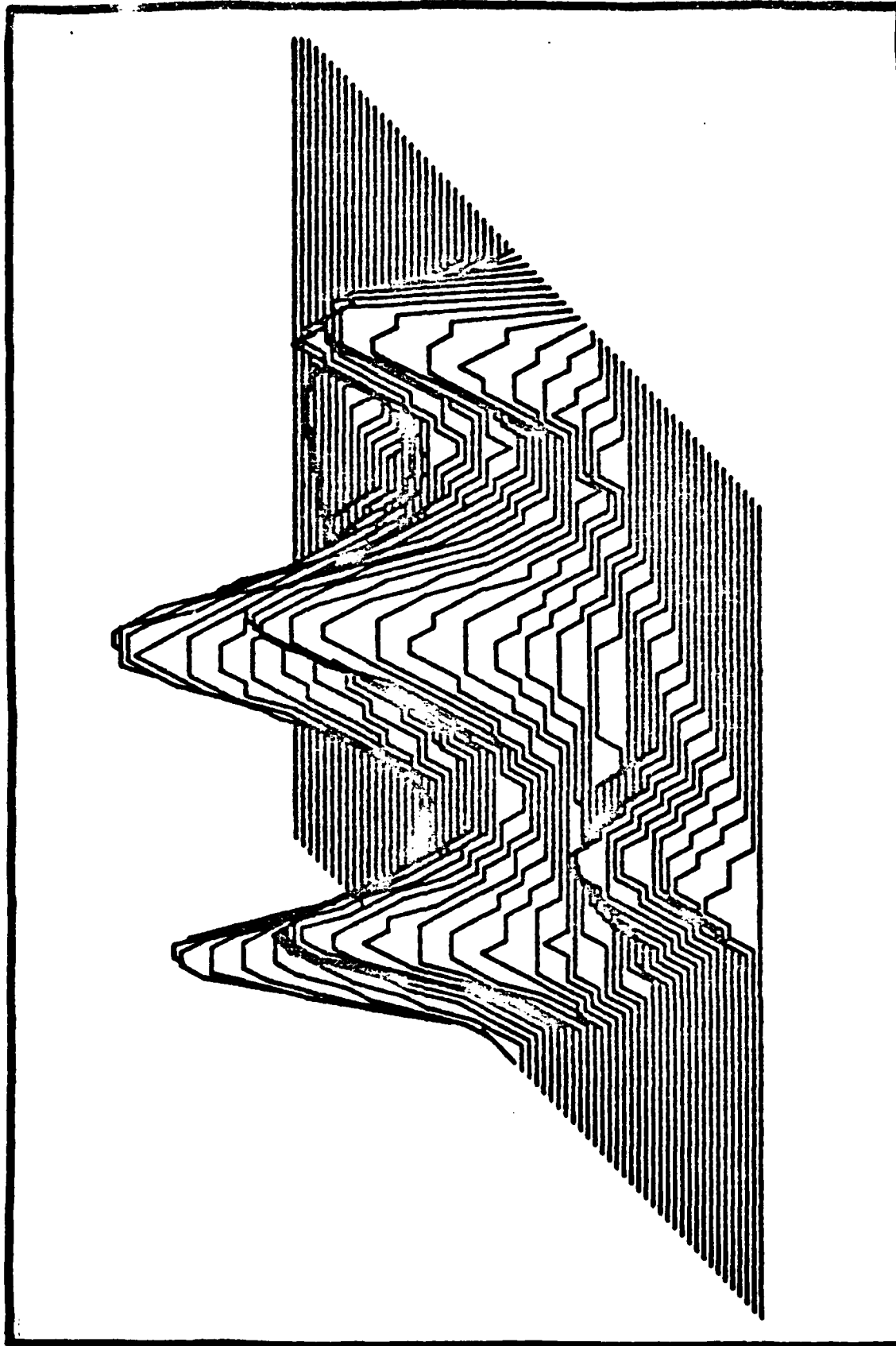


Fig. B.12. Minimum Cross-Entropy Template 12 with Cross-Entropy = 0.51369



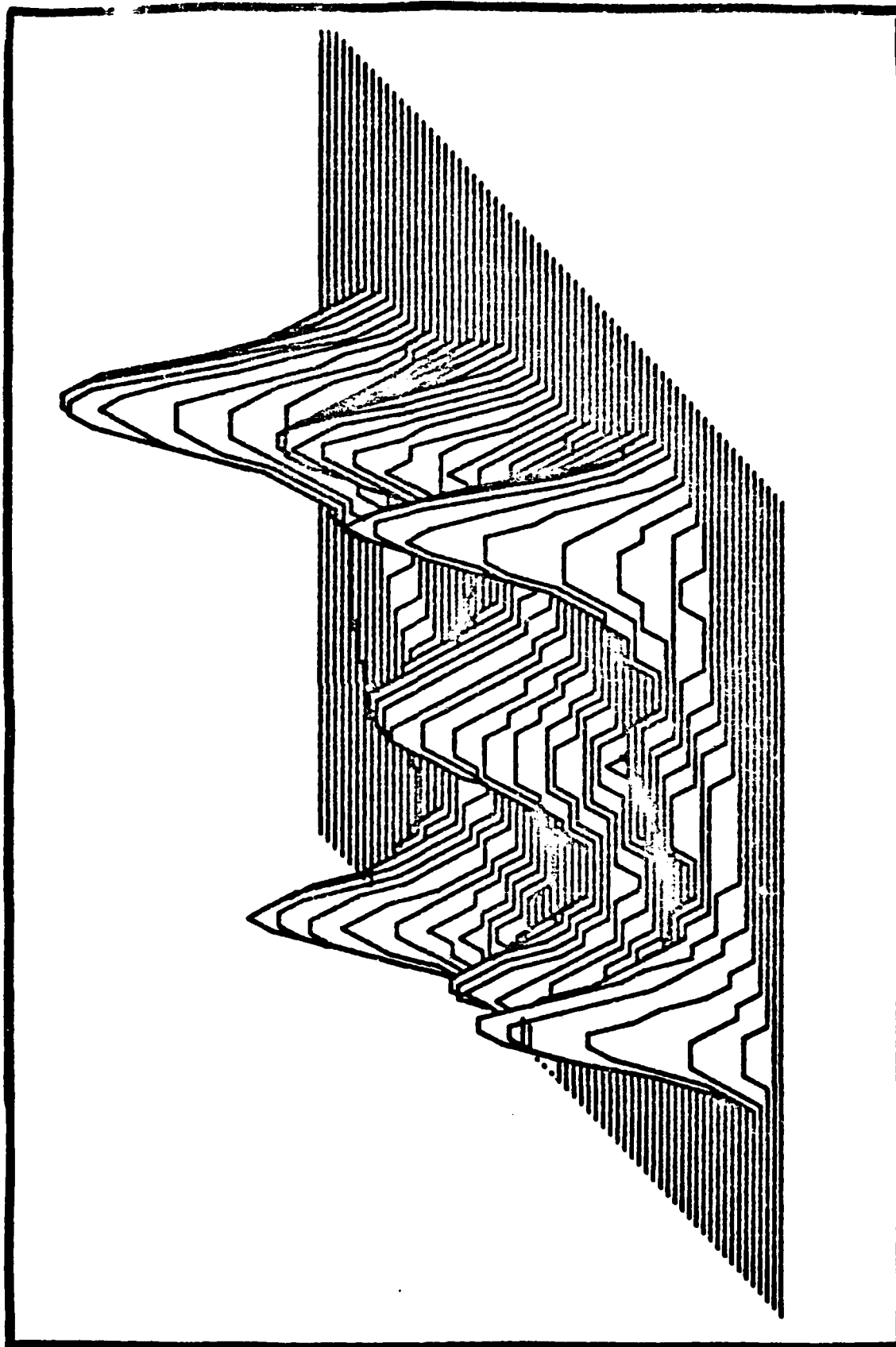


Fig. B.13. Minimum Cross-Entropy Template 13 with Cross-Entropy = 0.39674

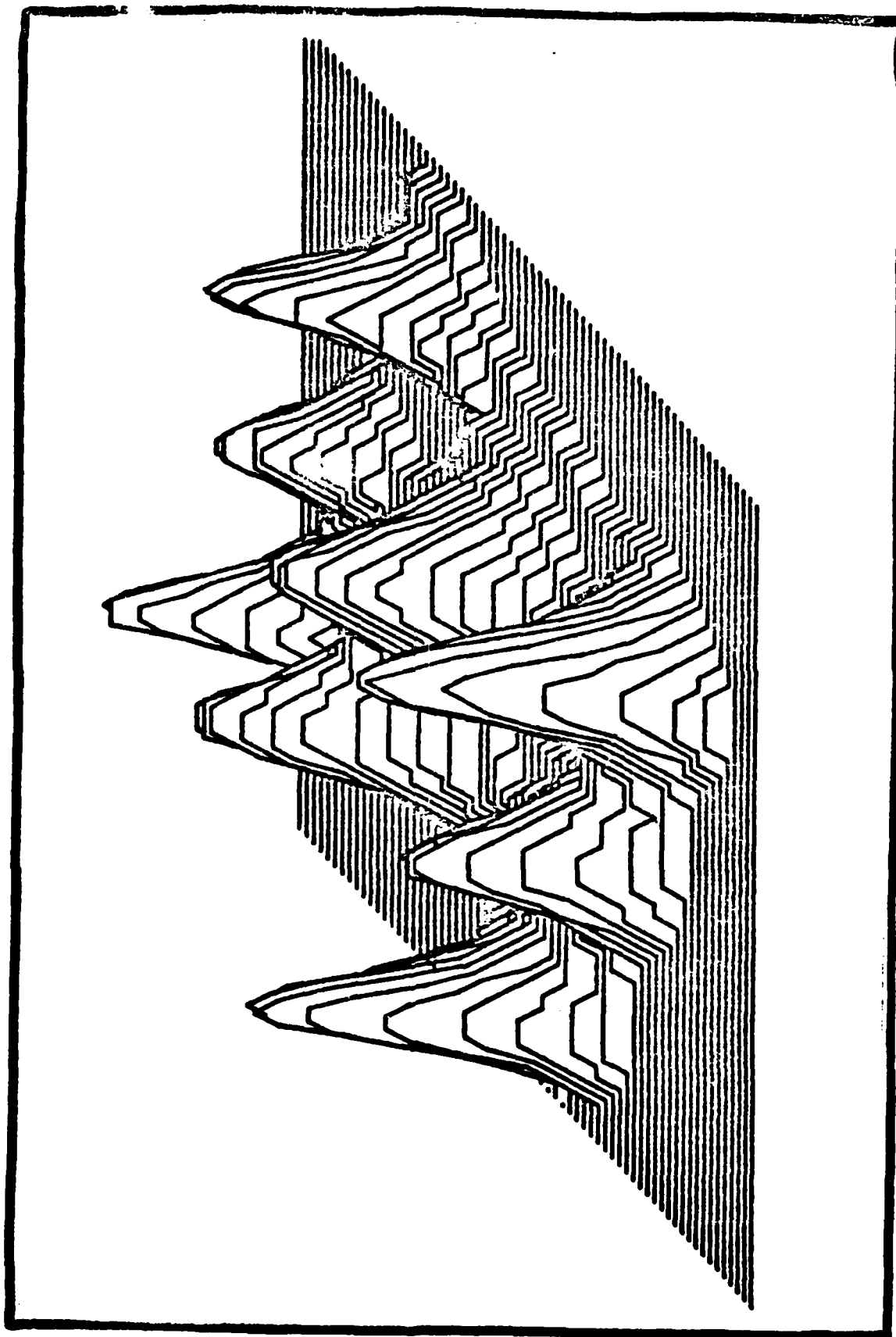


Fig. B.14. Minimum Cross-Entropy Template 14 with Cross-Entropy = 0.42777

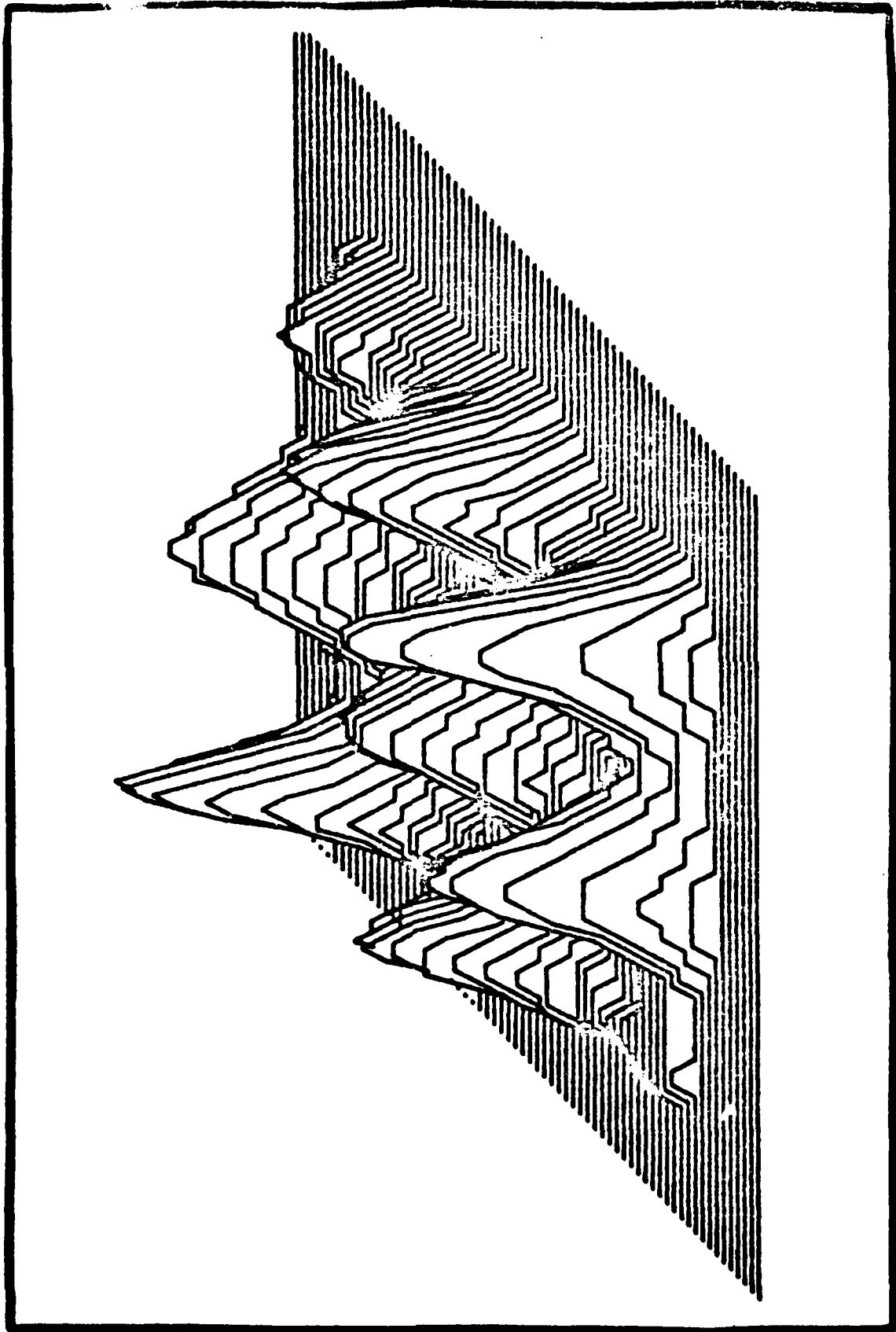


Fig. B.15. Minimum Cross-Entropy Template 15 with Cross-Entropy = 0.39324

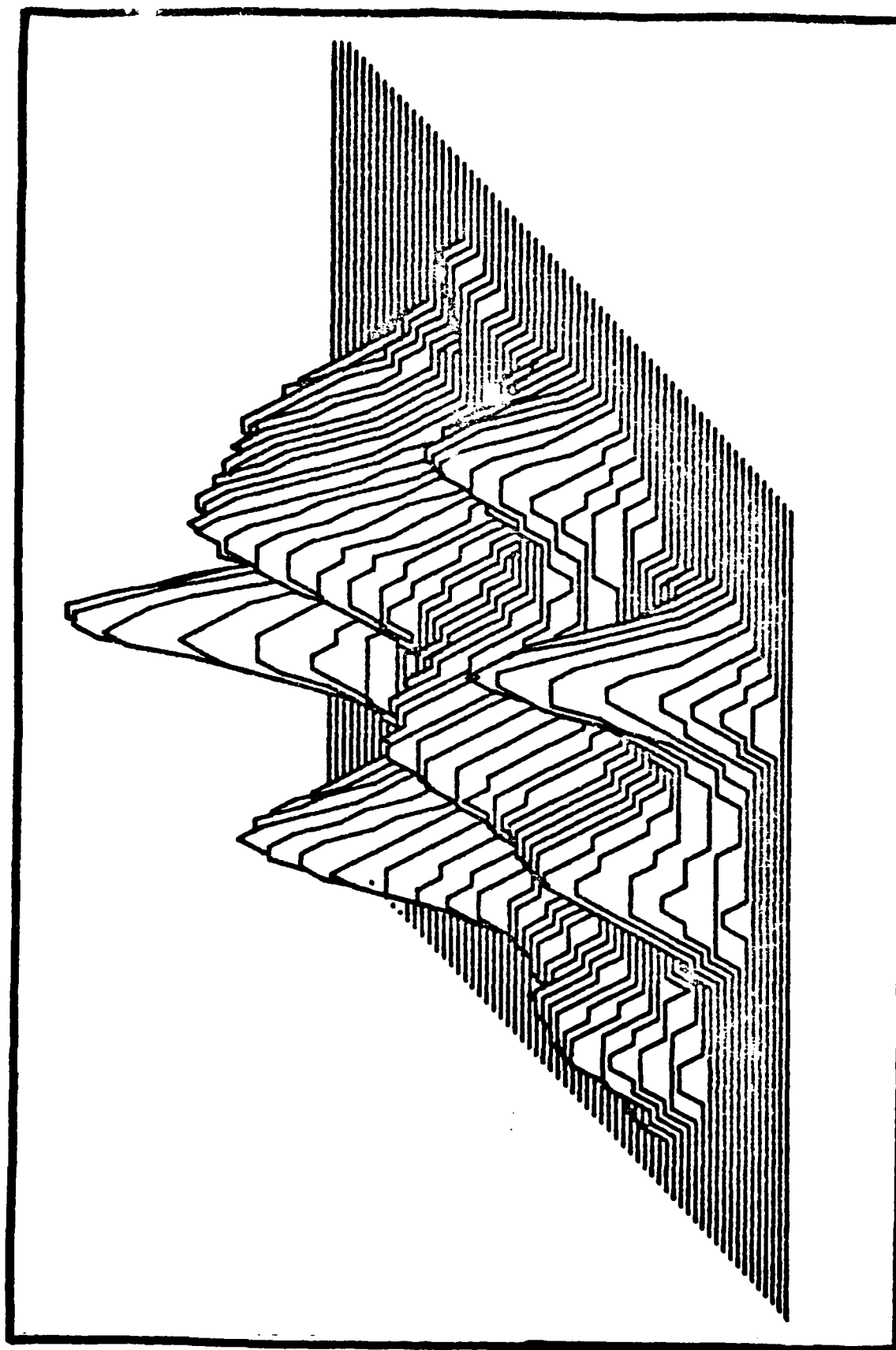


Fig. B.16. Minimum Cross-Entropy Template 16 with Cross-Entropy = 0.41373

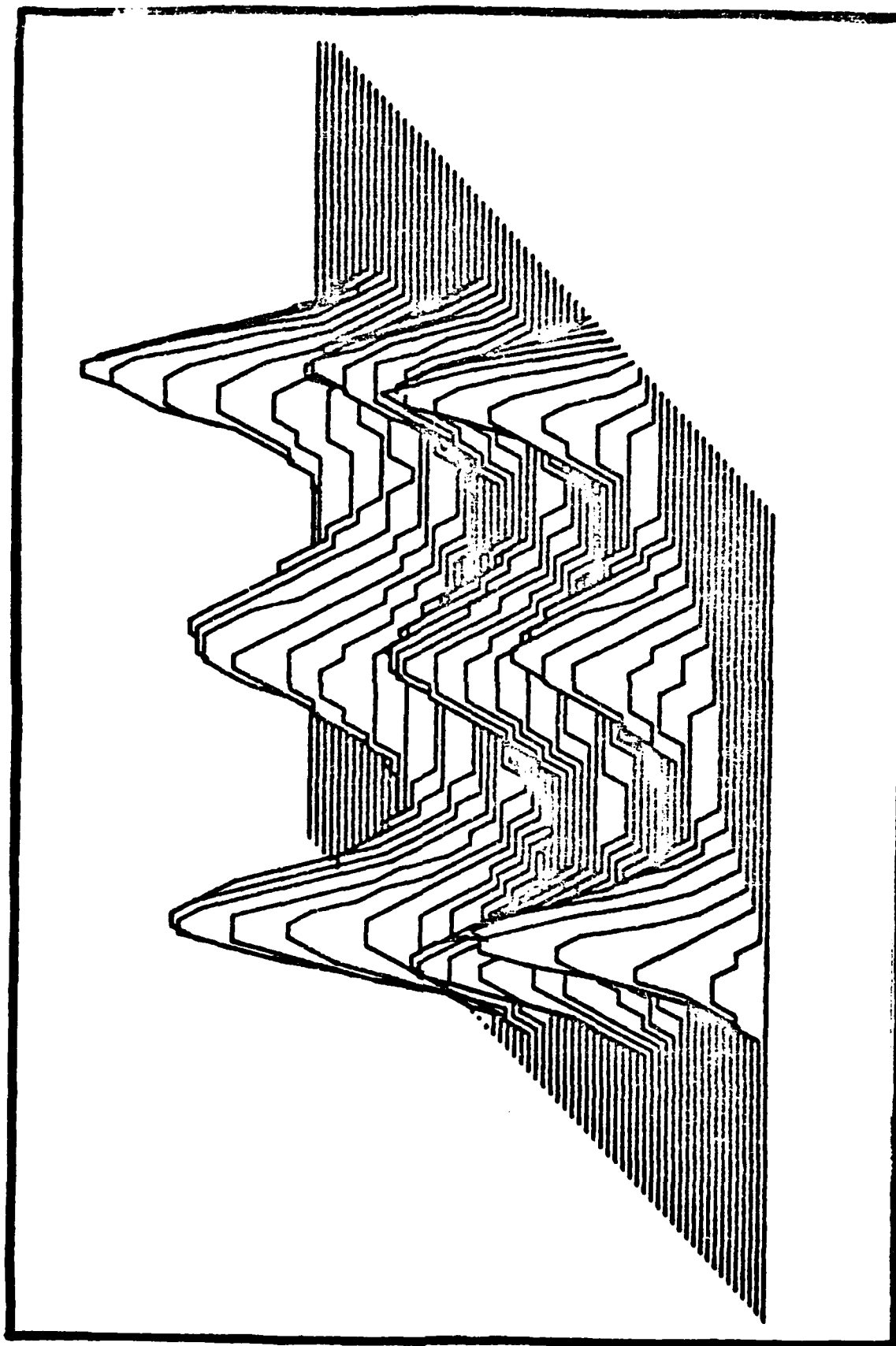


Fig. B.17. Minimum Cross-Entropy Template 17 with Cross-Entropy = 0.44381

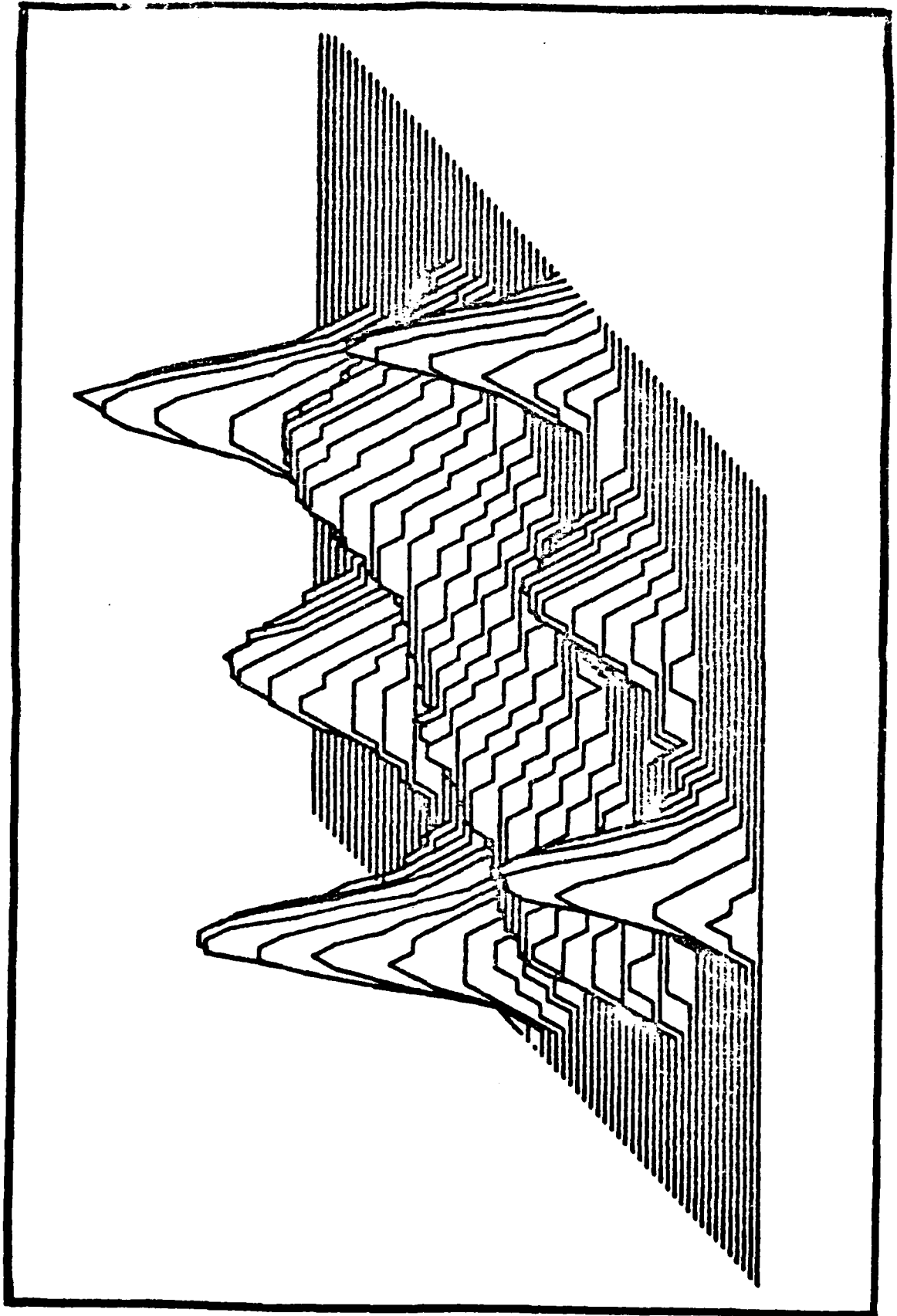


Fig. B.18. Minimum Cross-Entropy Template 18 with Cross-Entropy = 0.41452

### Appendix C. Template Photographs

The eighteen original template photographs are provided in this appendix. The photographs are sampled to produce the sixteen level gray scale perspective plots provided in Appendix A.



Template 1



Template 2

Fig. C.1. Template Alternative One



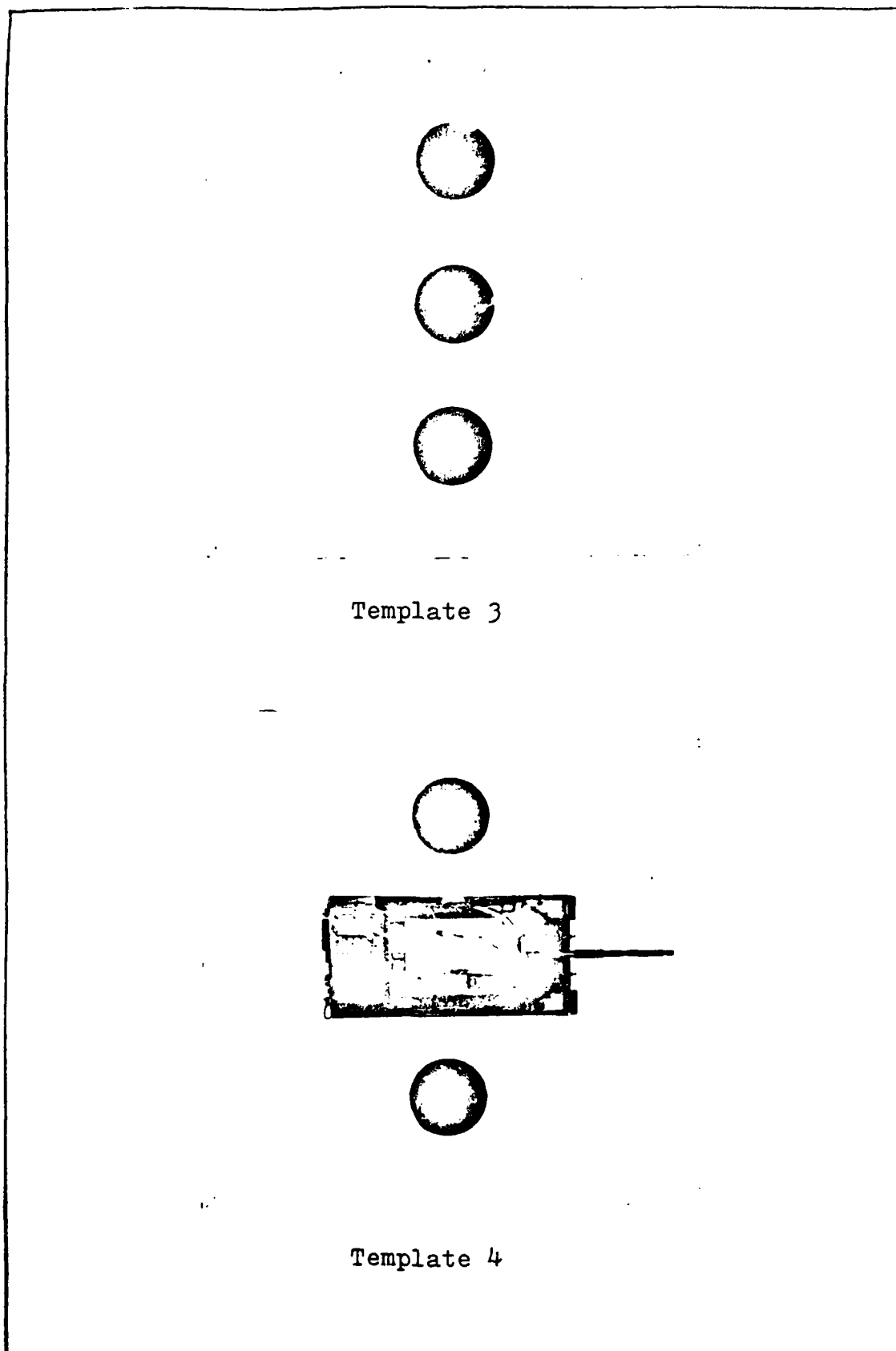


Fig. C.2. Template Alternative Two  
168

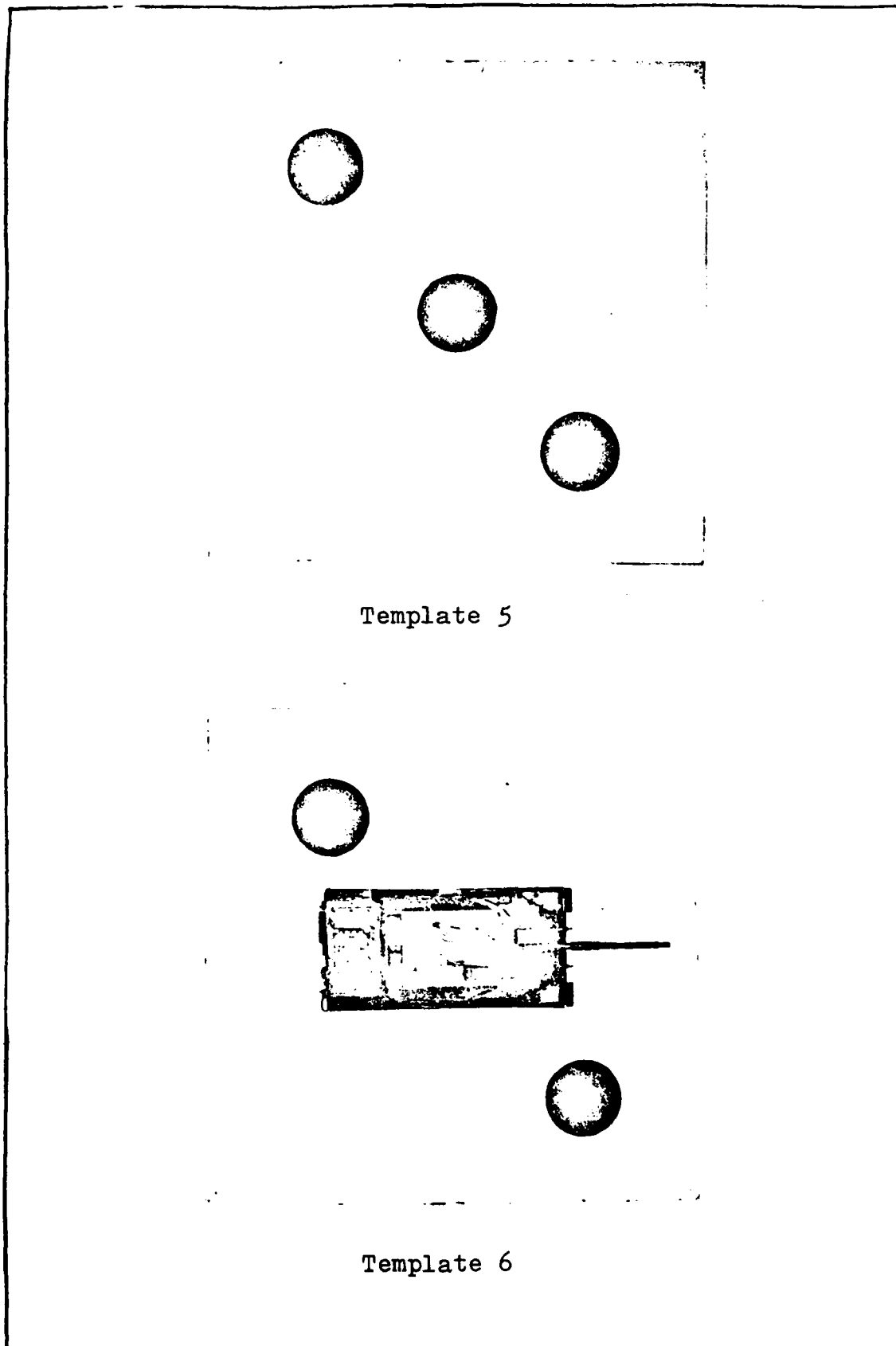
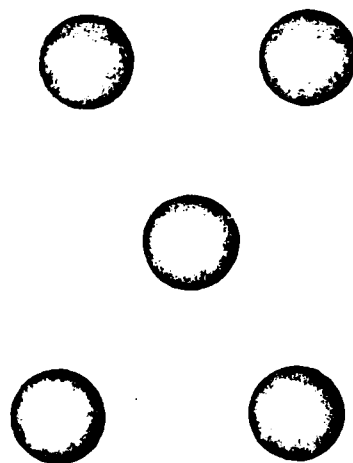
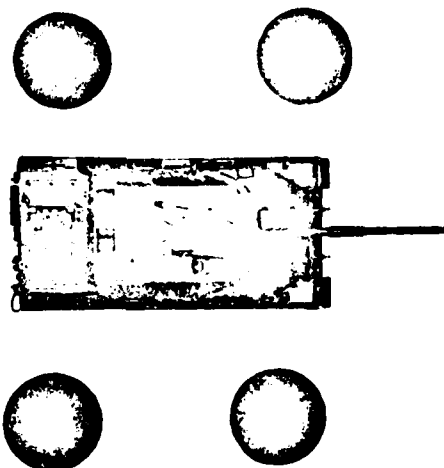


Fig. C.3. Template Alternative Three  
169



Template 7



Template 8

Fig. C.4 Template Alternative Four  
170

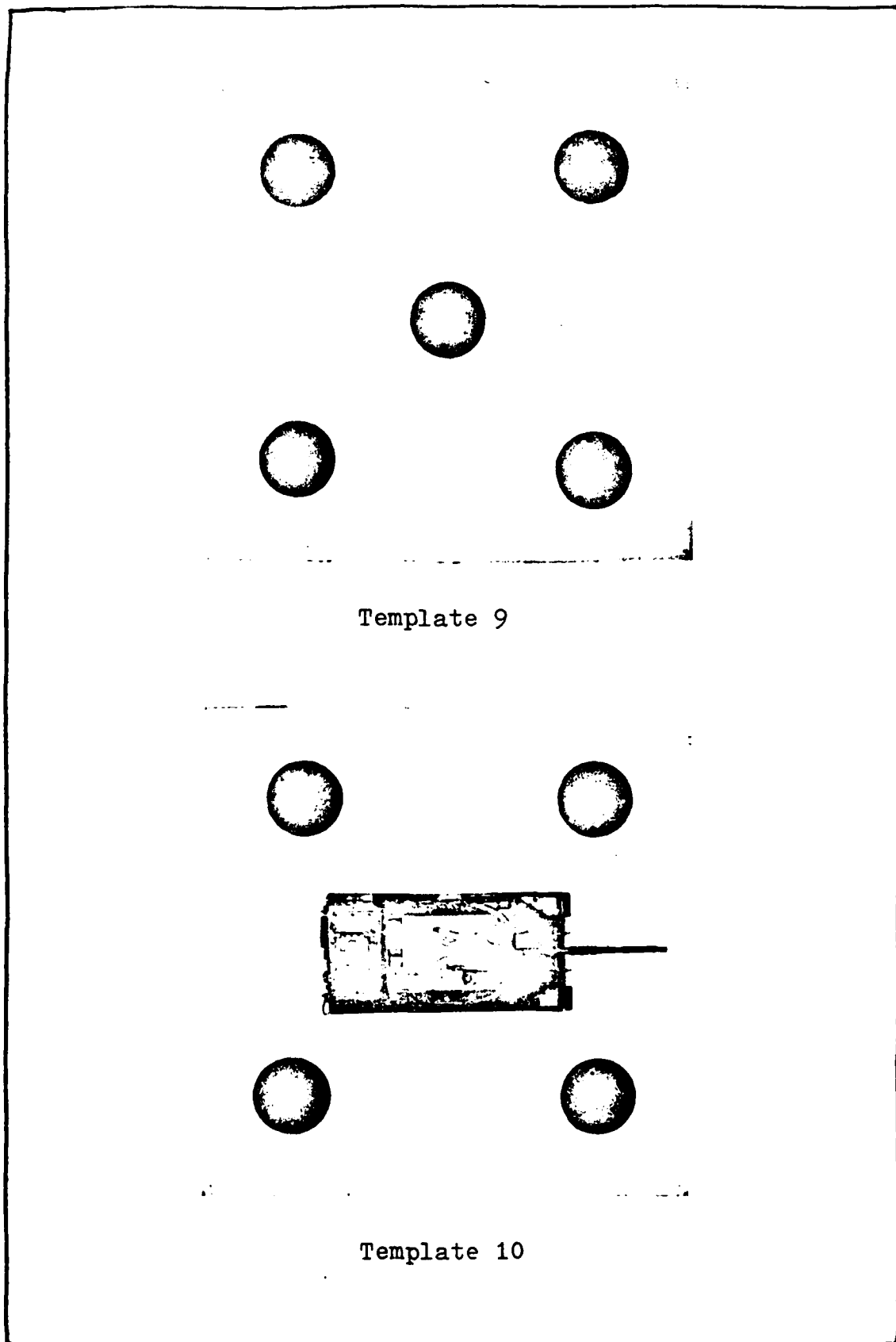
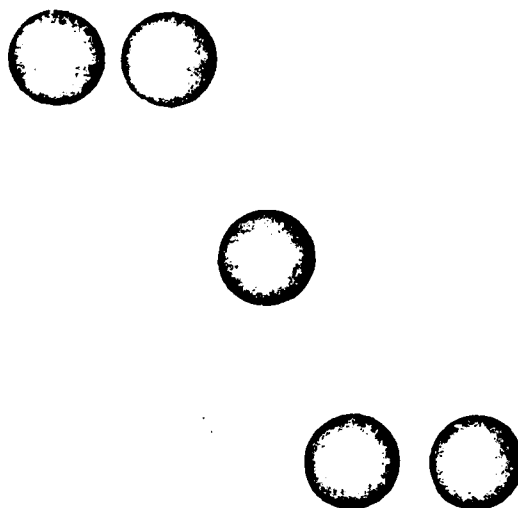
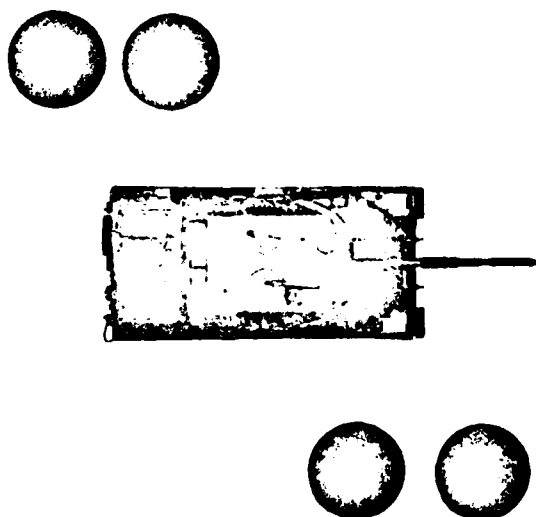


Fig. C.5. Template Alternative Five

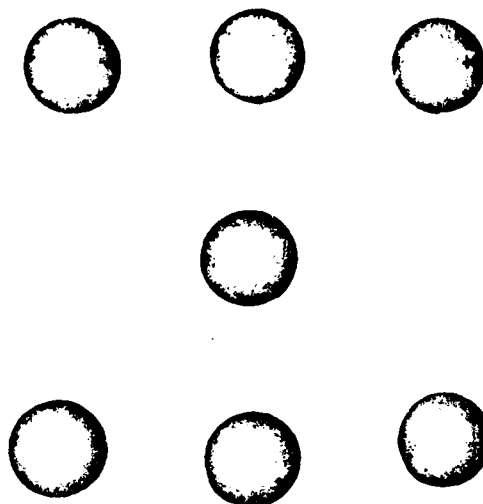


Template 11

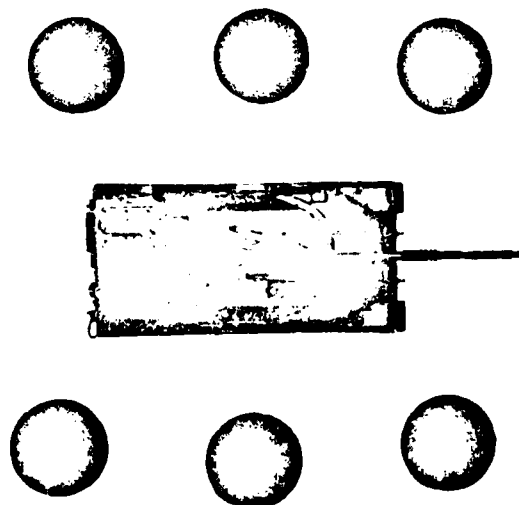


Template 12

Fig. C.6. Template Alternative Six

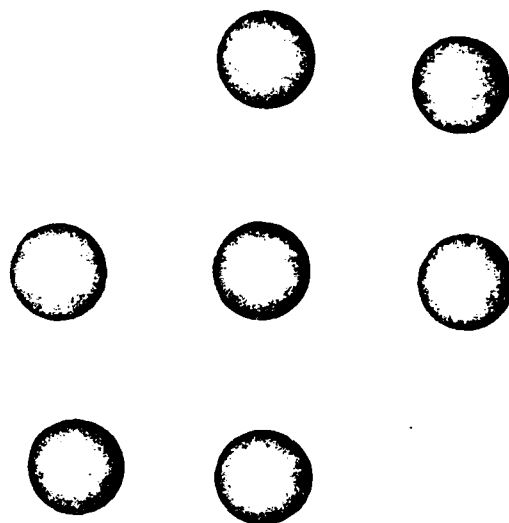


Template 13

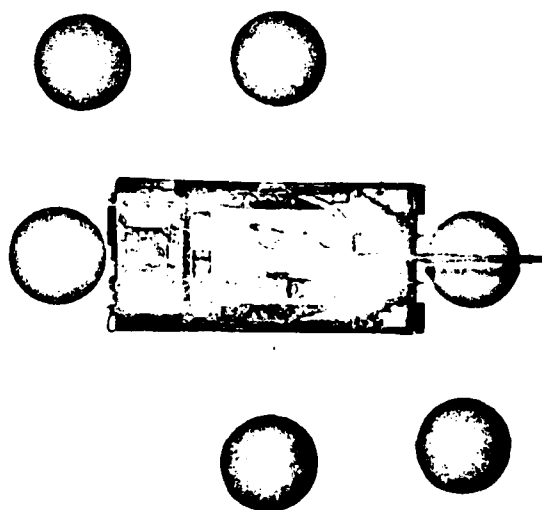


Template 14

Fig. C.7. Template Alternative Seven

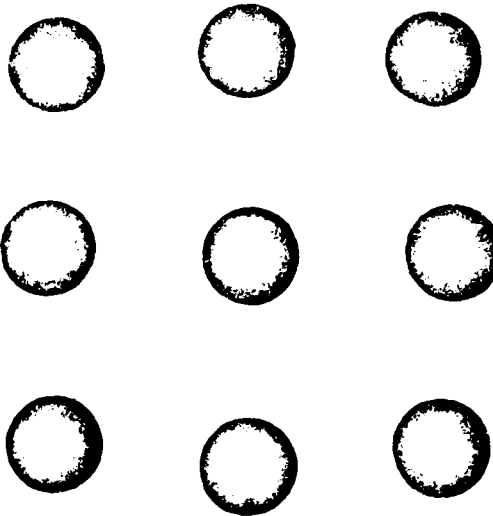


Template 15

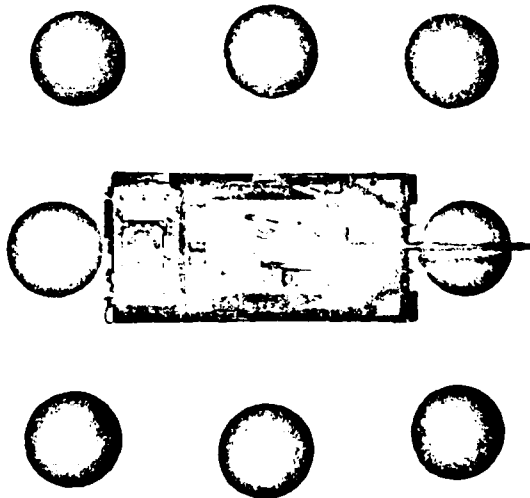


Template 16

Fig. C.8. Template Alternative Eight



Template 17



Template 18

Fig. C.9. Template Alternative Nine



### Vita

Frank Davis Tiller was born on June 7, 1947, in Portsmouth, Virginia. After graduating from Amherst County High School in Amherst, Virginia, he attended Virginia Polytechnic Institute from which he received the degree of Bachelor of Science in Electrical Engineering in 1970. He was employed by the General Electric Company in Hickory, North Carolina until December, 1970, when he entered Officer Training School and was commissioned in the USAF in April, 1971. Captain Tiller has held assignments in European Communications Area at Rhein Main AB, Germany and Tactical Communications Area, Langley AFB, Virginia. In January, 1978, Captain Tiller entered the Air Force Institute of Technology Graduate Electrical Engineering program at Wright-Patterson AFB, Ohio receiving the MSEE in December, 1979. Since receiving his master's degree, Captain Tiller has been enrolled in the Air Force Institute of Technology Doctor of Philosophy program working in the area of scene analysis.

Permanent Address: Route 1 Box 281 A

Madison Heights, Virginia 24572

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/DS/EE/82-1	2. GOVT ACCESSION NO. AD A124975	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Information Theoretic Detection of Objects Embedded in Cluttered Aerial Scenes		5. TYPE OF REPORT & PERIOD COVERED Ph.D. Dissertation
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Frank D. Tiller		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AFIT-EN) Wright-Patterson AFB, OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFWAL/AARM/3 Air Force Wright Aeronautical Laboratories Wright-Patterson AFB, OH 45433		12. REPORT DATE July 1982
		13. NUMBER OF PAGES 176
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES  Approved for public release; INW AFR 153-17. LYNN E. WOLFE Dean for Development Air Force Institute of Technology Wright-Patterson AFB, OH 45433 4 JAN 1983		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Image Processing Target Detection Probability Density Estimation Minimum Cross-Entropy  Maximum Entropy Information Theory		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  This dissertation presents the theoretical development and numerical implementation of a maximum cross-entropy target detection algorithm. The procedure is based on the solution of a non-linear constrained cross-entropy minimization problem and requires information in the form of raw image moments. The detection rule involves both preprocessing and real-time computations. The preprocessing requires the selection of a set of target templates and the solution of the constrained cross-entropy minimization problem.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified  
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20.

for the selected target templates. The real-time processing requires the computation of image moments and a set of dot product operations

An orthonormal set of "information functions" is developed and numerical methods of converting raw image moments into the expected values of the information functions are given. Numerical techniques for image moment computation and a solution scheme for the nonlinear set of constraints are developed and implemented. The theoretical development of the detection algorithm is given starting from a set of consistency axioms. The expected performance is analyzed and factors determining performance presented. The procedure is applied to a test set of 100 images and the detection algorithm error probability is projected and related to the salient performance determining factors.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

**END**

**FILMED**

**3-83**

**DTIC**